

Tilburg University

Semiparametric inference for non-LAN models

Zhou, Bo

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Zhou, B. (2017). *Semiparametric inference for non-LAN models*. [Doctoral Thesis, Tilburg University]. CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Semiparametric Inference for non-LAN Models

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 6 december 2017 om 10.00 uur door

Bo Zhou

geboren op 4 januari 1989 te Qinhuangdao, China.

Promotor: Prof. dr. Bas J.M. Werker

Copromotores: Dr. Ramon van den Akker

Dr. I. Gaia Becheri

Overige Commissie Leden: Dr. Otilia Boldea

Prof. dr. H. Peter Boswijk

Dr. Feike C. Drost

Prof. dr. John H.J. Einmahl

Acknowledgments

This dissertation was carried out at the Department of Econometrics & OR of Tilburg University, where I spent six colourful years of my life there. Six years is definitely not a short period, but when I am looking back, I feel it passed by so quickly. Probably it is because I enjoyed the time here so much, according to Einstein's theory of relativity. Now it comes to the final part of my dissertation and the subsequent Ph.D., I would like to take the chance to thank the people who have supported me in one way or another during this journey. It is not a fair task to put my heartfelt appreciation to them into a few words, but I will try my best.

First of all, I would like to express my sincere gratitude to my supervisor and my mentor, Prof. Bas Werker. He is a reliable and professional supervisor. His trustful advices and right amount of small pushes made me walk so far along this tough road and never fall down. He is also a happy, light-hearted person. I really enjoyed our meetings which always start with his "how's life?". Most importantly, he affects me permanently by setting a great example as a researcher pursuing truth. It is my luck to have him as my advisor.

I am also deeply grateful to my two co-supervisors, Dr. Ramon van den Akker and Dr. Gaia Becheri. When I entered the field of theoretical econometrics as a first year Ph.D. student, I knew almost nothing. It was Ramon who helped me go over the van den Varrrt's book, explained the details and the intuition behind it. My beginning would have been more difficult without his patient guidance, and I benefited a lot from his brilliant

ideas. Gaia is warm and sincere. I appreciate very much for her insightful questions during the meetings as well as kind relaxing chat to ease my nerves before my presentation.

It is a great honor for me to have Dr. Otilia Boldea, Prof. Peter Boswijk, Dr. Feike Drost, and Prof. John Einmahl as my Ph.D. committee. Their comments and suggestions during my pre-defence were very helpful and led to this improved final version of my dissertation. Moreover, I am also in debt for their unselfish assistance during my Ph.D life and the tough period of the job market.

My sincere thanks also goes to Prof. Per Mykland for offering the opportunity to visit the University of Chicago. It was a great time and a valuable experience, and thanks to his kind hospitality during my visit there.

I would like to acknowledge the Department of Econometrics & OR in Tilburg University. I benefit a lot from its excellent working environment and scientific research atmosphere. I am very thankful to Heidi, Lenie, Korine, and Anja for their thoughtful services.

I am thankful to my classmates, colleagues and friends: Alaa Abi Morshed, Hailong Bao, Nicola Ciccarelli, Zhuojiong Gan, Víctor Hugo Gonzalez, Yi He, Hong Li, Bas van Heiningen, Renata Rabovic, Mario Rothfelder, Chen Sun, Xiaoyu Wang, Yilong Xu, Yuxin Yao, Yifan Yu, Yifan Zhang, Nan Zhao, and many others. We had a great time together, and I hope our friendship will continue regardless of time and space.

I would like to thank to my parents for all their unconditional love, understanding and encouragement. They have always been supportive of my decisions. Finally, thank you my beloved girlfriend, Yan Xu. You brought me to this unexpected but pleasant journey, and we went throughout it side-by-side. Thank you for appearing in my life.

Bo Zhou

October, 2017

Contents

Acknowledgments	iii
1 Introduction	1
2 Semiparametrically Optimal Hybrid Rank Tests for Unit Roots	7
2.1 Introduction	8
2.2 The model	11
2.3 The power envelope for invariant tests	13
2.3.1 Preliminaries	13
2.3.2 A structural representation of the limit experiment . .	17
2.3.3 The limit experiment: invariance and power envelope	20
2.3.4 The asymptotic power envelope for asymptotically in- variant tests	25
2.4 Semiparametrically optimal hybrid rank tests	26
2.4.1 The hybrid rank tests based on a reference density . .	28
2.4.2 The approximate hybrid rank tests	33
2.5 Monte Carlo study	36
2.5.1 Large-sample performance	37
2.5.2 Small-sample performance	40
2.6 Discussions of possible extensions	43
2.6.1 Some simulation results	45
2.7 Conclusion	46
2.8 Appendix: Proofs	47

2.8.1	Preliminaries	47
2.8.2	Main proofs	49
3	A New Semiparametric Approach for Nonstandard Econo-	
	metric Problems	59
3.1	Introduction	60
3.2	The New Approach	65
3.2.1	Model Setup	65
3.2.2	Limit Likelihood Ratios	70
3.2.3	Structural Limit Experiment	71
3.2.4	Maximal Invariant and Semiparametric Efficiency Bounds	72
3.2.5	Semiparametric Efficient Inference with Rank Statistics	76
3.2.6	Some Discussions	78
3.3	Multivariate Rank Statistic	80
3.3.1	Standardized Error ν_t	80
3.3.2	Rank-based Score and Its Limit Behaviour	81
3.3.3	Construction of Rank-based Statistical Inference . . .	83
3.3.4	Optimality with Independent ν_t	86
3.4	Cointegration	87
3.4.1	The Model	87
3.4.2	Semiparametric Power Envelope	88
3.4.3	Rank-based Tests	91
3.4.4	Monte Carlo Simulations	92
3.4.5	Extensions	95
3.5	Weak Instrument	100
3.5.1	Model	100
3.5.2	Semiparametric Bound with Given π	101
3.5.3	Semiparametric Efficient Tests with Known f	103
3.5.4	Rank-based Efficient Tests	106
3.5.5	Monte Carlo Simulations	107
3.6	Conclusion	110
3.7	Appendix: Proofs	111

3.8	Appendix: Conditional Expectation Calculations	118
3.8.1	Linear Regression Case	118
3.8.2	Cointegration Case	120
3.8.3	Weak Instrument Case	120
4	Semiparametric Optimal Testing with Highly Persistent Pre-	
	dictors	123
4.1	Introduction	124
4.2	Model	126
4.2.1	Preliminaries	128
4.3	Eliminating μ and f by invariance	132
4.3.1	A Structural Representation of the Limit Experiment	133
4.3.2	Maximal Invariant	137
4.3.3	Semiparametric Power Envelope	138
4.3.4	Rank-based asymptotically invariant statistics	140
4.4	Eliminating the nuisance parameter γ by ALFD	144
4.4.1	The Approximately Least Favorable Distribution (ALFD)	
	Approach	145
4.4.2	Putting it all together	148
4.5	A Monte Carlo Study	150
4.6	Conclusions	154
4.7	Appendix: Proofs	158
4.7.1	Auxiliaries	158
4.7.2	Main Proofs	159
4.8	Appendix: Switching Tests to Standard Case	166

Chapter 1

Introduction

This dissertation focuses on developing semiparametric efficiency bounds and conducting semiparametric efficient inference for nonstandard econometric problems. These problems are nonstandard in the sense that their limit experiments are not of the standard *locally asymptotically normal* (LAN) form. Examples are the unit root testing problem, the cointegration rank testing problem, hypothesis testing problems with weak instruments, and the problem of predicting stock return with a highly persistent predictor, etc. The traditional *least favorable parametric submodel* (LFPS) method based on projecting the scores of the parameters of interest onto the tangent space generated by the nuisance parameters cannot be easily extended to these problems. Therefore, some new investigations into this direction need to be conducted.

The first essay in Chapter 2 can be seen as the first step. The problem is to test the unit root hypothesis in a univariate AR(1) model, in which the innovation density is treated as an infinite-dimensional nuisance parameter. Jansson (2008) applies the traditional LFPS approach and develops the semiparametric power upper bounds of both (asymptotically) invariant and similar tests. Then he shows that the bound for asymptotically invariant tests to be equal to the semiparametric power envelope by providing a feasible test attaining it, while the one for asymptotically similar tests is not

attainable. This essay follows in the line of Jansson's paper. We rederive the semiparametric power envelope for all asymptotically invariant tests. However, we make the invariance structures explicit, by using a new approach based on a nonparametric modeling of density and a structural version of the limit experiment. This structure has the advantage that it shows more clearly why we should employ the invariance restriction rather than the unbiasedness or similarity restriction. Moreover, invariance restrictions (and the corresponding Brownian bridges in the limit) naturally lead to the (partial) use of rank-based statistics. Therefore, we propose a new class of unit root tests based on the ranks of the increments of the observations, their average, and an assumed reference density for the innovations. We name these "Hybrid Rank Tests" (HRTs). The HRTs are semiparametric in the sense that they are valid, i.e., have the correct asymptotic size, irrespective of the true innovation density. For correctly specified reference density, our test is point-optimal and nearly efficient. For arbitrary reference density, we establish a Chernoff-Savage type result, i.e., our test performs as well as commonly used tests under Gaussian innovations but has improved power under other, e.g., fat-tailed or skewed, innovation distributions. We also propose a simplified version of our test that exhibits the same properties, where the Chernoff-Savage type result is restricted to Gaussian reference densities and demonstrated by simulation results.

As a subsequent step, in my second essay in Chapter 3, we provide a general framework to exploit invariance structures in semiparametric models where the likelihood ratios admit the *locally asymptotically Brownian functional* (LABF) form (see Jeganathan (1995)). Here we can regard LAN and *locally asymptotically mixed normal* (LAMN) models as special cases of LABF models. Unlike the traditional LFPS way that begins with some parametric submodels embedding the true unknown model (and finds the least-favorable one), in this new approach, we start with an explicit nonparametric modeling of the innovation density. In this model, we employ an orthonormal basis in a suitable functional space as perturbation functions,

and an infinite-dimensional local parameter η to describe the deviation from the true density. Then, using standard techniques, we obtain the LABF likelihood ratio under mild conditions. As this likelihood ratio can also be regarded as a Radon-Nikodym derivative, an application of Girsanov's Theorem leads to a structural version of the limit experiment. We call this the "structural limit experiment".

This structural limit experiment is key to my approach of exploiting invariance structures. In the structural limit experiment driven by LABF type likelihood ratios, we observe an infinite-dimensional process. It is a Brownian motion under the null hypothesis, while under the alternative hypothesis, it takes the form of an Ornstein-Uhlenbeck (OU) process. The nuisance parameter η appears in the drift term. This feature suggests to impose an invariance restriction to eliminate η . To be specific, taking the bridge of a process removes the drift term and, as a consequence, the obtained bridge process is invariant w.r.t any drift transformation in the original process. Applying this operation to all the elements (of the infinite-dimensional observation process) that are affected by η , we can get an invariant sigma-field. Even more, we prove that it is *maximally invariant*. This result is meaningful in the sense that any invariant statistic must be a function of this maximal invariant. Consequently, by the Neyman-Pearson lemma and the Asymptotic Representation Theorem, the power of the test based on the maximal invariant's likelihood ratio provides an upper bound for the power of all asymptotically invariant tests. In standard LAN models, we find that this new approach leads to the same efficiency bounds as the traditional LFPS method. Therefore, by this new approach, we extend the concept of *efficient score function* to all LABF cases.

Furthermore, the structural limit experiment and its invariance structures also suggest semiparametrically efficient inference procedures. We show in the essay that the LABF efficient score contains a stochastic integral of the form $\int_0^1 M(u, h) dB_{\ell_f}(u)$, where B_{ℓ_f} is the Brownian bridge of a given Brownian motion W_{ℓ_f} . A simple derivation shows that this term also

equals $\int_0^1 \widetilde{M}(u, h) dW_{\ell_f}(u)$ with $\widetilde{M}(u, h) = M(u, h) - \int_0^1 M(u, h) du$. These two expressions are reminiscent of two strands in the statistical literature respectively: inference based on a nonparametric estimate \hat{f} of the unknown density f , and inference based on rank statistics. The former \hat{f} -based inference is standard for both univariate case and multivariate case. However, the latter rank-based inference is difficult to extend to the multivariate case. This is because vectors in \mathbb{R}^d , $d \geq 2$, are not naturally ordered. One common approach in the existing literature is to assume that f is elliptical, which allows these vectors to be ordered by the Mahalanobis distance. The assumption of an elliptical density may be restrictive in many applications. To relax this assumption, we propose an inference procedure based on componentwise rank statistics, the score structure of the multivariate normal distribution, and d arbitrary marginal reference densities. Simulation results show substantial efficiency gains when the innovations are “far from” normal distributed, e.g., multivariate Student’s t_3 distributed.

We apply the above new approach to two nonstandard problems: testing the cointegration rank and testing hypotheses with weak instruments. In the cointegration application, the semiparametric power envelope is developed first, and then a component-wise rank-based test is proposed. In the weak instrument application, we first reveal the limiting insights of the AR test by Anderson, and Rubin (1949), the LM test by Kleibergen (2002), and the CLR test by Moreira (2003), using the structural limit experiment in the Gaussian case. Then we derive the semiparametric optimal version for these tests, and propose rank-based versions of these three tests.

The third essay in Chapter 4 deals with the problem of testing predictability of, say, stock returns when the predictor variable is highly persistent. This problem is also nonstandard in the sense that it is of the LABF type (not LAN or LAMN) and, moreover, the nuisance local autocorrelation parameter of the predictor c will appear under the null hypothesis when the innovations of the returns and those of the predictors are correlated. The literature mainly focuses on the latter issue (eliminating the nuisance param-

eter c) and proposes several ways to handle it. References are, for instance, Campbell and Yogo (2005) which uses the Bonferroni method, Jansson and Moreira (2006) which proposes to impose a conditionality restriction on the exponential family form of the Gaussian log-likelihood ratio, and the recent Elliott, Müller, and Watson (2015) approach based on a numerically calculated *approximately least favorable distribution (ALFD)*. However, all the methods above are based on a Gaussianity assumption. For non-Gaussian distributions, the quest for semiparametric efficiency still needs investigation. For this purpose, using our new semiparametric approach, we first develop the semiparametric power envelope for the case where c is known. As a natural subsequent step, we propose an efficient test statistic based on the componentwise ranks of the innovations and freely chosen marginal reference densities. To eliminate the nuisance parameter c , we subsequently employ the ALFD approach of Elliott, Müller, and Watson (2015), since it enjoys superior size and power properties compared to other methods. Simulation results show that my test gains considerable efficiency when the innovation density is multivariate Student's t_3 distributed.

Chapter 2

Semiparametrically Optimal Hybrid Rank Tests for Unit Roots

[Based on joint work with Ramon van den Akker and Bas Werker
Semiparametrically Optimal Hybrid Rank Tests for Unit Roots.]

Abstract. We propose a new class of unit root tests that exploits invariance properties in the Locally Asymptotically Brownian Functional limit experiment associated to the standard unit root model. The invariance structures naturally suggest tests that are based on the ranks of the increments of the observations, their average, and an assumed reference density for the innovations. The tests are semiparametric in the sense that they are valid, i.e., have the correct (asymptotic) size, irrespective of the true innovation density. For correctly specified reference density, our test is point-optimal and nearly efficient. For arbitrary reference density, we establish a Chernoff-Savage type result, i.e., our test performs as well as commonly used tests under Gaussian innovations but has improved power under other, e.g., fat-tailed or skewed, innovation distributions. We also propose a simplified

version of our test that exhibits the same properties, however the Chernoff-Savage type result is restricted to Gaussian reference densities and can only be demonstrated by simulations.

Key words. Unit root test, semiparametric power envelope, limit experiment, LABF, maximal invariant, rank statistic.

2.1 Introduction

The recent monographs of Patterson (2011, 2012) and Choi (2015) provide an overview of the literature on unit roots tests. This literature traces back to White (1958) and includes seminal papers as Dickey and Fuller (1979, 1981), Phillips (1987), Phillips and Perron (1988), and Elliott, Rothenberg, and Stock (1996). The present paper fits into the stream of literature that focuses on “optimal” testing for unit roots. Important early contributions here are Dufour and King (1991), Saikkonen and Luukkonen (1993), and Elliott, Rothenberg, and Stock (1996). The latter paper derives the asymptotic power envelope for unit root testing in models with Gaussian innovations. Rothenberg and Stock (1997) and Jansson (2008) consider subsequently the non-Gaussian case.

The present paper considers testing for unit roots in a semiparametric setting. Following earlier literature, we focus on a simple AR(1) model driven by i.i.d. innovations whose distribution is considered a nuisance parameter. Apart from some smoothness and the existence of relevant moments, no assumptions are imposed on this distribution. From earlier work it is known that the unit root model leads to Locally Asymptotically Brownian Functional (LABF) limit experiments (in the Le Cam sense). As a consequence, no uniformly most powerful test exists (even in case the innovation distribution would be known) – see also Elliott, Rothenberg, and Stock (1996). In the semiparametric case the limit experiment becomes even more difficult, precisely due to the infinite-dimensional nuisance parameter. Jansson (2008) derives the semiparametric power envelope by mimicking

ideas that hold for Locally Asymptotically Normal (LAN) models. However, the proposed test needs a full nonparametric score function estimator which complicates its implementation. Our optimal test only requires a nonparametric estimation of a real-valued cross-information factor.

The main contribution of this manuscript is twofold. First, we provide a new derivation of the semiparametric asymptotic power envelope for unit root tests (Section 2.3). This derivation is build upon invariance structures embedded in the semiparametric unit root model. To be precise, we use a “structural” description of the LABF limit experiment (Section 2.3.2), obtained from Girsanov’s theorem. This limit experiment corresponds to observing an infinitely-dimensional Ornstein-Uhlenbeck process (on the time interval $[0, 1]$). The unknown innovation density in the semiparametric unit root model, takes the form of an unknown drift function in the limit experiment. Within this limit experiment, we subsequently (Section 2.3.3) derive the maximal invariant, i.e., a reduction of the data which is invariant with respect to the nuisance parameters (that is, the unknown drift in the limiting Ornstein-Uhlenbeck experiment). It turns out that this maximal invariant takes a rather simple form (all components, but one, of the multivariate process have to be replaced by their associated bridges). The power envelope for invariant tests in the limit experiment then follows from the Neyman-Pearson lemma. An application of the Asymptotic Representation Theorem subsequently yields the local asymptotic power envelope (Theorem 2.3.2). We note that our analysis of invariance structures in the LABF experiment is also of independent interest and could, for example, be exploited in the analysis of optimal inference for cointegration or predictive regression models. Moreover, it also gives an alternative interpretation of the test proposed in Elliott, Rothenberg, and Stock (1996) — the ERS test — as it is also based on an invariant, though not the maximal one.

As a second contribution, we provide a new class of easy-to-implement unit root tests that are semiparametrically optimal in the sense that their asymptotic power curve is tangent to the semiparametric power envelope

(Section 2.4.1). The form of the maximal invariant developed before suggests how to construct such tests based on the ranks of the increments of the observations, the average of these increments, and an assumed reference density g . These tests are semiparametric in the sense that the reference density need not equal the true innovation density, while they still provide the correct asymptotic size. This reference density is not restricted to be Gaussian, which it generally is in more classical QMLE results. When the reference density is correctly specified (i.e, g equals to the true density f), the asymptotic power curve of our test is tangent to the semiparametric power envelope, and this in turn gives the optimality property.¹ Following Elliott, Rothenberg, and Stock (1996) we also discuss the selection of a fixed alternative that yields a “nearly efficient” test, i.e., one for which the asymptotic local power function is uniformly close to the semiparametric power envelope.² Our tests, despite the absence of a LAN structure, satisfy a Chernoff and Savage (1958) type result (Corollary 2.4.1): with any reference density our test outperforms, at any true density, its classical counterpart which in this case, is the ERS test. We provide, in Section 4.2, an even simpler alternative class of tests. Both classes of tests coincide for correctly specified reference density and, thus, share the same optimality properties. In case of misspecified reference density, the alternative class still seems to enjoy the Chernoff-Savage type property, though only for Gaussian reference density. This is in line with the traditional Chernoff-Savage results for Locally Asymptotically Normal models.

The remainder of this paper is organized as follows. Section 2.2 introduces the model assumptions and some notation. Next, Section 2.3 con-

¹A feasible solution for the oracle requirement $g = f$ for optimality is to use a non-parametrically estimated density \hat{f} . A detailed discussion of this point is provided in Section 2.6.

²Here the concept of “nearly efficient” is borrowed from Elliott, Rothenberg, and Stock (1996) and Jansson and Nielsen (2012). It is based on simulation results rather than rigorous mathematical proof, since there is no uniformly most powerful tests in this unit root testing problem.

tains the analysis of the limit experiment. In particular we study invariance properties in the limit experiment leading to our new derivation of the semi-parametric power envelope. The class of hybrid rank tests we propose is introduced in Section 2.4. Section 4.5 provides the results of a Monte Carlo study and Section 3.6 contains a discussion of possible extensions of our results. All proofs are organized in Section 2.8.

2.2 The model

We consider observations Y_1, \dots, Y_T generated from the classical component specification

$$Y_t = \mu + X_t, \quad t \in \mathbb{N}, \quad (2.2.1)$$

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \quad (2.2.2)$$

where $X_0 = 0$ and the innovations $\{\varepsilon_t\}$ form an i.i.d. sequence with density f . We impose the following assumptions on this innovation density.

Assumption 2.2.1.

- (a) The density f is absolutely continuous with a.e. derivative f' , i.e. for all $a < b$ we have

$$f(b) - f(a) = \int_a^b f'(e) de.$$

- (b) $E_f[\varepsilon_t] = \int e f(e) de = 0$ and $\sigma_f^2 = \text{Var}_f[\varepsilon_t] < \infty$.

- (c) The standardized Fisher-information for location,

$$J_f = \sigma_f^2 \int \phi_f^2(e) f(e) de,$$

where $\phi_f(e) = -(f'/f)(e)$ is the *location score*, is finite.

- (d) The density f is positive, i.e., $f > 0$.

Let \mathcal{F} denote the set of densities satisfying Assumption 2.2.1.

The imposed smoothness assumptions (a) on f are mild and standard. The finite variance assumption (b) is important to our asymptotic results as it is essential to the weak convergence, to a Brownian motion, of the partial-sum process generated by the innovations.³ The zero mean assumption in (b) excludes a deterministic trend in the model. Such a trend leads to an entirely different asymptotic analysis, see Hallin, Van den Akker, and Werker (2011). The Fisher information J_f in (c) has been standardized by premultiplying with the variance σ_f^2 , so that it becomes scale invariant (i.e., invariant w.r.t. σ_f). In other words, J_f only depends on the shape of the density f and not on its variance σ_f^2 . The positivity of the density f in (d) is mainly made for notational convenience. The assumption on the initial condition, $X_0 = 0$, is less innocent then it may appear. Indeed, it is known, see Müller and Elliott (2003) and Elliott and Müller (2006), that, even asymptotically, the initial condition can contain non-negligible statistical information.

The main goal of this paper is to develop tests, with optimality features, for the semiparametric unit root hypothesis

$$H_0 : \rho = 1, (\mu \in \mathbb{R}, f \in \mathcal{F}) \text{ versus } H_a : \rho < 1, (\mu \in \mathbb{R}, f \in \mathcal{F}),$$

i.e., apart from Assumption 2.2.1, no further structure is imposed on f and the intercept μ is also treated as a nuisance parameter. It is well-known, and goes back to Phillips (1987), Chan and Wei (1988) and Phillips and Perron (1988), that the contiguity rate for the unit root testing problem, i.e., the fastest convergence rate at which it is possible to distinguish (with non-trivial power) the unit root $\rho = 1$ from a stationary alternative $\rho < 1$, is

³Let us already mention that, although not allowed for in our theoretical results, we will also assess the finite-sample performances of the proposed tests (Section 4.5) for innovation distributions with infinite variance. For tests specifically developed for such cases we refer to Hasan (2001), Ahn, Fotopoulos, and He (2003), and Callegari, Cappuccio, and Lubian (2003).

given by T^{-1} . Therefore, in order to compare performances of tests with this proper rate of convergence, we reparametrize the autoregression parameter ρ into its local-to-unity form, i.e.,

$$\rho = \rho_h^{(T)} = 1 + \frac{h}{T}, \quad (2.2.3)$$

and we can rewrite our hypothesis of interest as

$$H_0 : h = 0, (\mu \in \mathbb{R}, f \in \mathcal{F}) \text{ versus } H_a : h < 0, (\mu \in \mathbb{R}, f \in \mathcal{F}).$$

In the following section, we derive the (asymptotic) power envelope of tests that are (asymptotically) invariant with respect to the nuisance parameters μ and f . Section 2.4 is subsequently devoted to tests, depending on a reference density g that can be freely chosen, that are point optimal with respect to this power envelope and proves the Chernoff-Savage result.

2.3 The power envelope for invariant tests

This section first introduces some notations and preliminaries (Section 2.3.1). Next, we will derive the limit experiment (in the Le Cam sense) corresponding to the component unit root model (2.2.1)-(2.2.2) and provide a “structural” representation of this limit experiment (Section 2.3.2). In Section 2.3.3 we discuss, exploiting this structural representation, a natural invariance restriction, to be imposed on tests for the unit root hypothesis with respect to the infinite-dimensional nuisance parameter associated to the innovation density. We derive the maximal invariant and obtain from this the power envelope for invariant tests in the limit experiment.

2.3.1 Preliminaries

We first discuss a convenient parametrization of perturbations to the innovation density which we use to deal with the semiparametric nature of the testing problem. These perturbations follow the standard approach of local alternatives in (semiparametric) models commonly used in experiments that

are Locally Asymptotically Normal (LAN). We will see that, with respect to the innovation density f alone, the model is actually LAN; compare also Remark 2.3.1 below. Moreover, we introduce some partial sum processes that we need in the sequel, as well as their Brownian limits.

Perturbations to the innovation density

To describe the local perturbations to the density f , we need the separable Hilbert space

$$L_2^{0,f} = L_2^{0,f}(\mathbb{R}, \mathcal{B}) = \left\{ b \in L_2^f(\mathbb{R}, \mathcal{B}) \mid \int b(e)f(e)de = 0, \int eb(e)f(e)de = 0 \right\},$$

where $L_2^f(\mathbb{R}, \mathcal{B})$ denotes the space of Lebesgue-measurable functions $b : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $\int b^2(e)f(e)de < \infty$. Because of the separability, there exists a countable orthonormal basis b_k , $k \in \mathbb{N}$, of $L_2^{0,f}$. This basis can be chosen such that $b_k \in C_{2,b}(\mathbb{R})$, for all k , i.e., each b_k is bounded and two times continuously differentiable with bounded derivatives. Hence each function $b \in L_2^{0,f}$ can be written as $b = \sum_{k=1}^{\infty} \eta_k b_k$, for some $(\eta_k)_{k \in \mathbb{N}} \in \ell_2 = \{(x_k)_{k \in \mathbb{N}} \mid \sum_{k=1}^{\infty} x_k^2 < \infty\}$. Besides the sequence space ℓ_2 we also need the sequence space c_{00} which is defined as the set of sequences with finite support, i.e.,

$$c_{00} = \left\{ (x_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{k=1}^{\infty} 1\{x_k \neq 0\} < \infty \right\}.$$

Of course, c_{00} is a dense subspace of ℓ_2 . For $b_k \in L_2^{0,f}$ with $\text{Var}_f b_k(\varepsilon) = 1$, $\eta \in c_{00}$ we now introduce the following perturbation to the density f :

$$f_{\eta}^{(T)}(e) = f(e) \left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k(e) \right), \quad e \in \mathbb{R}. \quad (2.3.1)$$

The rate $T^{-1/2}$ is already indicative of the standard LAN behavior of the nuisance parameter f as will formally follow from Proposition 2.3.2 below. The following proposition shows that these perturbations are valid in the sense that they satisfy the conditions on the innovation density that we imposed throughout on the model (Assumption 2.2.1).

Proposition 2.3.1. *Let f satisfy Assumption 2.2.1 and suppose $\eta \in c_{00}$. Then there exists $T' \in \mathbb{N}$ such that for all $T \geq T'$ we have $f_\eta^{(T)} \in \mathcal{F}$.*

Remark 2.3.1. In semiparametric statistics one typically parametrizes perturbations (paths in semiparametric parlor) to a density by a so-called “non-parametric” score function $h \in L_2^{0,f}$, i.e., a perturbation takes the form $f(e)k(T^{-1/2}h(e)) \approx f(e)(1 + T^{-1/2}h(e))$ for a suitable function k ; see, for example, Bickel et al. (1998) for details. By using the basis b_k , $k \in \mathbb{N}$, we instead tackle all such perturbations via the infinite-dimensional nuisance parameter η . Of course, one would need to use ℓ_2 as parameter space to “generate” all score functions h . We instead restrict to c_{00} which ensures (4.2.8) to be a density (for large T). For our purposes this restriction will be without cost. Intuitively, this is since c_{00} is a dense subspace of ℓ_2 (so if a property is “sufficiently continuous” one only needs to establish it on c_{00} because it extends to the closure).

Partial sum processes

To describe the limit experiment in Section 2.3.2, we introduce some partial sum processes and their limits. These results are fairly classical but, for completeness, precise statements are organized in Lemma A.1 in the supplementary material.

As usual, Δ denotes differencing, i.e., $\Delta Y_t = Y_t - Y_{t-1}$. Define, for $s \in [0, 1]$,

$$\begin{aligned} W_\varepsilon^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \frac{\Delta Y_t}{\sigma_f}, \\ W_{\phi_f}^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \sigma_f \phi_f(\Delta Y_t), \quad f \in \mathcal{F}, \\ W_{b_k}^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} b_k(\Delta Y_t), \quad k \in \mathbb{N}. \end{aligned}$$

The rationale of our notation is that we have $\Delta Y_t = \varepsilon_t$, for $t \geq 2$, under the null hypothesis of a unit root. Also note that the sums start at $t = 2$, so the partial sum processes are (maximally) invariant with respect to the intercept μ . Using Assumption 2.2.1 we find, under the null hypothesis, weak convergence⁴ of $W_\varepsilon^{(T)}$, $W_{\phi_f}^{(T)}$, and $W_{b_k}^{(T)}$ to Brownian motions that we denote by W_ε , W_{ϕ_f} , and W_{b_k} , respectively. These limiting Brownian motions W_ε , W_{ϕ_f} , and W_{b_k} are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0})$. Let us already mention that we will introduce a collection of probability measures $\mathbb{P}_{h,\eta}$ representing the limit experiment, in Section 2.3.2. We use the notational convention that probability measures related to the limit experiment (i.e., to the Brownian motions) are denoted by \mathbb{P} , while probability measures related to the finite-sample unit root model will be denoted by $\mathbb{P}^{(T)}$.

As ε and $b_k(\varepsilon)$ are orthogonal for each k , we find that W_ε and W_{b_k} , $k \in \mathbb{N}$, are all mutually independent. Moreover,

$$\text{Var}_{0,0}[W_\varepsilon(1)] = 1 \quad \text{and} \quad \text{Var}_{0,0}[W_{b_k}(1)] = 1.$$

As $\phi_f(\varepsilon)$ is the score of the location model, it is well known (see, for example, Bickel et al. (1998)) that we have (under Assumption 2.2.1) $\mathbb{E}_f[\phi_f(\varepsilon)] = 0$ and $\mathbb{E}_f[\varepsilon\phi_f(\varepsilon)] = 1$. Consequently, again because ε and $b_k(\varepsilon)$ are orthogonal for each k , we can decompose $\sigma_f\phi_f(\varepsilon) = \sigma_f^{-1}\varepsilon + \sum_{k=1}^{\infty} J_{f,k}b_k(\varepsilon)$, with coefficients $J_{f,k} = \sigma_f\mathbb{E}_f[b_k(\varepsilon)\phi_f(\varepsilon)]$. This establishes, for $f \in \mathcal{F}$,

$$W_{\phi_f} = W_\varepsilon + \sum_{k=1}^{\infty} J_{f,k}W_{b_k}. \quad (2.3.2)$$

Moreover, we have, for $k \in \mathbb{N}$

$$\text{Cov}_{0,0}(W_{\phi_f}(1), W_\varepsilon(1)) = 1, \quad \text{Cov}_{0,0}(W_{\phi_f}(1), W_{b_k}(1)) = J_{f,k} \quad (2.3.3)$$

and

$$\text{Var}_{0,0}[W_{\phi_f}(1)] = J_f = 1 + \sum_{k=1}^{\infty} J_{f,k}^2. \quad (2.3.4)$$

⁴All weak convergences in this paper are in product spaces of $D[0,1]$ with the uniform topology.

We remark that integrals like $\int_0^1 W_\varepsilon^{(T)}(s-)dW_{\phi_f}^{(T)}(s)$ can be shown to converge weakly to the associated stochastic integral with the limiting Brownian motions, i.e., to $\int_0^1 W_\varepsilon(s)dW_{\phi_f}(s)$. Weak convergence of integrals like $\int_0^1 (W_\varepsilon^{(T)}(s-))^2 ds$ follows from an application of the continuous mapping theorem. Again, details are provided in Section 2.8.

2.3.2 A structural representation of the limit experiment

The results in the previous section are needed to study the asymptotic behavior of log-likelihood ratios. These in turn determine the limit experiment, which we use to study asymptotically optimal procedures invariant with respect to the nuisance parameters f and μ . Thus, fix $f \in \mathcal{F}$ and $\mu \in \mathbb{R}$. Let, for $h \in \mathbb{R}$ and $\eta \in c_{00}$, $P_{h,\eta;\mu,f}^{(T)}$ denote the law of Y_1, \dots, Y_T under (2.2.1)-(2.2.2) with autoregression parameter ρ given by (2.2.3) and innovation density (4.2.8). The following proposition shows that the semiparametric unit root model is of the Locally Asymptotically Brownian Functional (LABF) type introduced in Jeganathan (1995).

Proposition 2.3.2. *Let $\mu \in \mathbb{R}$, $f \in \mathcal{F}$, $\eta \in c_{00}$, and $h \in \mathbb{R}$.*

(i) *Then we have, under $P_{0,0;\mu,f}^{(T)}$,*

$$\begin{aligned} \log \frac{dP_{h,\eta;\mu,f}^{(T)}}{dP_{0,0;\mu,f}^{(T)}} &= \log \frac{f_\eta^{(T)}(Y_1 - \mu)}{f(Y_1 - \mu)} + \sum_{t=2}^T \log \frac{f_\eta^{(T)}(\Delta Y_t - \frac{h}{T}(Y_{t-1} - \mu))}{f(\Delta Y_t)} \\ &= h\Delta_f^{(T)} + \sum_{k=1}^{\infty} \eta_k \Delta_{b_k}^{(T)} - \frac{1}{2} \mathcal{I}_f^{(T)}(h, \eta) + o_P(1), 1 \end{aligned}$$

where the central-sequence $\Delta^{(T)} = (\Delta_f^{(T)}, \Delta_b^{(T)})$, with $\Delta_b^{(T)} = (\Delta_{b_k}^{(T)})_{k \in \mathbb{N}}$, is given by

$$\begin{aligned} \Delta_f^{(T)} &= \int_0^1 W_\varepsilon^{(T)}(s-)dW_{\phi_f}^{(T)}(s) = \frac{1}{T} \sum_{t=2}^T (Y_{t-1} - Y_1)\phi_f(\Delta Y_t), \\ \Delta_{b_k}^{(T)} &= W_{b_k}^{(T)}(1) = \frac{1}{\sqrt{T}} \sum_{t=2}^T b_k(\Delta Y_t), \quad k \in \mathbb{N}, \end{aligned}$$

and

$$\begin{aligned}\mathcal{I}_f^{(T)}(h, \eta) &= h^2 J_f \int_0^1 (W_\varepsilon^{(T)}(s-))^2 ds + \|\eta\|_2^2 + 2h \int_0^1 W_\varepsilon^{(T)}(s-) ds \sum_{k=1}^{\infty} \eta_k J_{f,k} \\ &= h^2 J_f \frac{1}{T^2} \sum_{t=2}^T \frac{(Y_{t-1} - Y_1)^2}{\sigma_f^2} + \|\eta\|_2^2 + 2h \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{(Y_{t-1} - Y_1)}{\sigma_f} \sum_{k=1}^{\infty} \eta_k J_{f,k}.\end{aligned}$$

(ii) Moreover, with $\Delta_f = \int_0^1 W_\varepsilon(s) dW_{\phi_f}(s)$ and $\Delta_{b_k} = W_{b_k}(1)$, $k \in \mathbb{N}$, we have, still under $\mathbb{P}_{0,0;\mu,f}^{(T)}$ and as $T \rightarrow \infty$,

$$\frac{d\mathbb{P}_{h,\eta;\mu,f}^{(T)}}{d\mathbb{P}_{0,0;\mu,f}^{(T)}} \Rightarrow \exp \left(h\Delta_f + \sum_{k=1}^{\infty} \eta_k \Delta_{b_k} - \frac{1}{2} \mathcal{I}_f(h, \eta) \right), \quad (2.3.5)$$

where

$$\mathcal{I}_f(h, \eta) = h^2 J_f \int_0^1 (W_\varepsilon(s))^2 ds + \|\eta\|_2^2 + 2h \int_0^1 W_\varepsilon(s) ds \sum_{k=1}^{\infty} \eta_k J_{f,k}.$$

(iii) For all $h \in \mathbb{R}$ and $\eta \in c_{00}$ the right-hand side of (4.3.1) has unit expectation under $\mathbb{P}_{0,0}$.

The proof of (i) follows by an application of Proposition 1 in Hallin et al. (2015) which provides generally applicable sufficient conditions for the quadratic expansion of log likelihood ratios. Of course, Part (ii) is not surprising and follows using the weak convergence of the partial sum processes to Brownian motions (and integrals involving the partial sum processes to stochastic integrals) discussed above. Finally, Part (iii) follows by verifying the Novikov condition. Detailed proofs are organized in Section 2.8.

Part (iii) of the proposition implies that we can introduce, for $h \in \mathbb{R}$ and $\eta \in c_{00}$, new probability measures $\mathbb{P}_{h,\eta}$ on the measurable space (Ω, \mathcal{F}) (on which the processes W_ε , W_{ϕ_f} , and W_{b_k} were defined) by their Radon-Nikodym derivatives with respect to $\mathbb{P}_{0,0}$:

$$\frac{d\mathbb{P}_{h,\eta}}{d\mathbb{P}_{0,0}} = \exp \left(h\Delta_f + \sum_{k=1}^{\infty} \eta_k \Delta_{b_k} - \frac{1}{2} \mathcal{I}_f(h, \eta) \right).$$

Proposition 2.3.2 then implies that the sequence of unit root experiments (each $T \in \mathbb{N}$ yields an experiment) weakly converges (in the Le Cam sense) to the experiment described by the probability measures $\mathbb{P}_{h,\eta}$. Formally, we define the sequence of experiments of interest by

$$\mathcal{E}^{(T)}(\mu, f) = \left(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), (\mathbb{P}_{h,\eta;\mu,f}^{(T)} \mid h \in \mathbb{R}, \eta \in c_{00}) \right), \quad T \in \mathbb{N},$$

and the limit experiment by, with \mathcal{B}_C the Borel σ -field on $C[0, 1]$,

$$\mathcal{E}(f) = \left(C[0, 1] \times C^{\mathbb{N}}[0, 1], \mathcal{B}_C \otimes (\otimes_{k=1}^{\infty} \mathcal{B}_C), (\mathbb{P}_{h,\eta} \mid h \in \mathbb{R}, \eta \in c_{00}) \right).$$

Corollary 2.3.1. *Let $\mu \in \mathbb{R}$ and $f \in \mathcal{F}$. Then the sequence of experiments $\mathcal{E}^{(T)}(\mu, f)$, $T \in \mathbb{N}$, converges (as $T \rightarrow \infty$) to the experiment $\mathcal{E}(f)$.*

The Asymptotic Representation Theorem (see, e.g., Chapter 9 in Van der Vaart (2000)) implies that for any statistic A_T which converges in distribution to the law $L_{h,\eta}$, under $\mathbb{P}_{h,\eta;\mu,f}^{(T)}$, there exists a (randomized) statistic A , defined on $\mathcal{E}(f)$, such that the law of A under $\mathbb{P}_{h,\eta}$ is given by $L_{h,\eta}$. This allows us to study (asymptotically) optimal inference: the “best” procedure in the limit experiment yields a bound for the sequence of experiments. If one is able to construct a statistic (for the sequence) that attains this bound, it follows that the bound is sharp and the statistic is called (asymptotically) optimal. This is precisely what we do: Section 2.3.3 establishes the bound and in Section 2.4 we introduce a statistic attaining it.

To obtain more insight in the limit experiment $\mathcal{E}(f)$ the following proposition, which follows by an application of Girsanov’s theorem, provides a “structural” description of the limit experiment.

Proposition 2.3.3. *Let $f \in \mathcal{F}$, $\eta \in c_{00}$, and $h \in \mathbb{R}$. Then the processes Z_ε and Z_{b_k} , $k \in \mathbb{N}$, defined by the starting values $Z_\varepsilon(0) = Z_{b_k}(0) = 0$ and the stochastic differential equations, for $s \in [0, 1]$,*

$$dZ_\varepsilon(s) = dW_\varepsilon(s) - hW_\varepsilon(s)ds,$$

$$dZ_{b_k}(s) = dW_{b_k}(s) - hJ_{f,k}W_\varepsilon(s)ds - \eta_k ds, \quad k \in \mathbb{N},$$

are Brownian motions under $\mathbb{P}_{h,\eta}$. Their joint law is that of $(W_\varepsilon, (W_{b_k})_{k \in \mathbb{N}})$ under $\mathbb{P}_{0,0}$.

Remark 2.3.2. Part (i) and (ii) Proposition 2.3.2 show that the parameter μ vanishes in the limit. More explicitly, in the proof of this proposition, we replace μ in the likelihood ratio term by Y_1 and then show that the difference term is $o_P(1)$. On the other hand, one could also localize the parameter μ as $\mu = \mu_d^{(T)} = \mu_0 + d$ (with rate T^0) as in Jansson (2008). As shown in that paper, the term associated to parameter d does not change with T and is independent to the other terms of the likelihood ratio. By the additively separable structure, we can treat parameter μ “as if” it is known. In either way, the inference for β would be invariant with respect to μ in the limit. Analogously, in the finite-sample experiment $\mathcal{E}^{(T)}(f)$, μ is eliminated (automatically) by using the increments ΔY_t , $t = 1, \dots, T$, which are (maximally) invariant with respect to any transformation on μ (see Section 2.4).

2.3.3 The limit experiment: invariance and power envelope

Using Proposition 2.3.3 we first discuss a natural invariance structure, with respect to the infinite-dimensional nuisance parameter η , for the limit experiment. We derive the maximal invariant and apply the Neyman-Pearson lemma to obtain the power envelope for invariant tests in the limit experiment. In Section 2.3.4 we then exploit the Asymptotic Representation Theorem to translate these results to obtain (asymptotically) optimal invariant test in the sequence of unit root models.

Consider the testing problem for the limit experiment $\mathcal{E}(f)$. We thus observe the processes W_ε and W_{b_k} , $k \in \mathbb{N}$, (continuously) on the time interval $[0, 1]$ from the model $(\mathbb{P}_{h,\eta} \mid h \in \mathbb{R}, \eta \in c_{00})$. We are interested in the power

envelope for testing the hypothesis

$$H_0 : h = 0, (\eta \in c_{00}) \text{ versus } H_a : h < 0, (\eta \in c_{00}). \quad (2.3.6)$$

We focus on test statistics that are invariant with respect to the value of the nuisance parameter η , i.e., these test statistics take the same value irrespective of the value of η . We now formalize this invariance structure.

Introduce, for $\eta \in c_{00}$, the transformation $g_\eta = (g_{\eta_k})_{k \in \mathbb{N}} : C^{\mathbb{N}}[0, 1] \rightarrow C^{\mathbb{N}}[0, 1]$ defined by, for $W \in C[0, 1]$,

$$g_{\eta_k} : [g_{\eta_k}(W)](s) = W(s) - \eta_k s, \quad s \in [0, 1], \quad (2.3.7)$$

i.e., g_{η_k} adds a drift $s \mapsto -\eta_k s$ to W . Proposition 2.3.3 implies that the law of $(W_\varepsilon, (g_{\eta_k}(W_{b_k}))_{k \in \mathbb{N}})$ under $\mathbb{P}_{h,0}$ is the same as the law of $(W_\varepsilon, (W_{b_k})_{k \in \mathbb{N}})$ under $\mathbb{P}_{h,\eta}$. Hence our testing problem (2.3.6) is invariant with respect to the transformations g_η . Therefore, following the invariance principle, it is natural to restrict attention to test statistics that are invariant with respect to these transformations as well, i.e., test statistics t that satisfy

$$t(W_\varepsilon, (g_{\eta_k}(W_{b_k}))_{k \in \mathbb{N}}) = t(W_\varepsilon, (W_{b_k})_{k \in \mathbb{N}}) \text{ for all } g_\eta, \eta \in c_{00}. \quad (2.3.8)$$

Given a process W let us define the associated *bridge process* by $B^W(s) = W(s) - sW(1)$. Now note that we have, for all $s \in [0, 1]$ and $k \in \mathbb{N}$,

$$\begin{aligned} B^{g_{\eta_k}(W)}(s) &= [g_{\eta_k}(W)](s) - s[g_{\eta_k}(W)](1) \\ &= W(s) - s\eta_k - s(W(1) - 1 \times \eta_k) \\ &= W(s) - sW(1) \\ &= B^W(s), \end{aligned}$$

i.e., taking the bridge of a process ensures invariance with respect to adding drifts to that process. Define the mapping M by $M(W_\varepsilon, (W_{b_k})_{k \in \mathbb{N}}) := (W_\varepsilon, (B_{b_k})_{k \in \mathbb{N}})$, with $B_{b_k} = B^{W_{b_k}}$. It follows that statistics that are measurable with respect to the σ -field,

$$\mathcal{M} = \sigma(M(W_\varepsilon, (W_{b_k})_{k \in \mathbb{N}})) = \sigma(W_\varepsilon, (B_{b_k})_{k \in \mathbb{N}}), \quad (2.3.9)$$

are invariant (with respect to g_η , $\eta \in c_{00}$). It is, however, not clear that we did not throw away too much data. Formally, we need \mathcal{M} to be *maximally invariant* which means that each invariant statistic is \mathcal{M} -measurable. The following theorem, which once more exploits the structural description of the limit experiment, shows that this indeed is the case.

Theorem 2.3.1. *The σ -field \mathcal{M} in (2.3.9) is maximally invariant for the group of transformations g_η , $\eta \in c_{00}$, in the experiment $\mathcal{E}(f)$.*

The above theorem implies that invariant inference must be based on \mathcal{M} . An application of the Neyman-Pearson lemma, using \mathcal{M} as observation, yields the power envelope for the class of invariant tests. To be precise, consider the likelihood ratios restricted to \mathcal{M} , which are given by

$$\frac{d\mathbb{P}_h^{\mathcal{M}}}{d\mathbb{P}_0^{\mathcal{M}}} = \mathbb{E}_0 \left[\frac{d\mathbb{P}_{h,\eta}}{d\mathbb{P}_{0,\eta}} \mid \mathcal{M} \right],$$

where the conditional expectation indeed does not depend on η precisely because of the invariance. To calculate this conditional expectation we first introduce $B_{\phi_f} = B^{W_{\phi_f}}$, i.e., the bridge process associated to W_{ϕ_f} defined in (2.3.2). Now we can decompose $\Delta_f = \int_0^1 W_\varepsilon(s) dW_{\phi_f}(s) = I + II$ with

$$\begin{aligned} I &= \int_0^1 W_\varepsilon(s) dB_{\phi_f}(s) + W_\varepsilon(1) \int_0^1 W_\varepsilon(s) ds, \\ II &= \left(\sum_{k=1}^{\infty} J_{f,k} W_{b_k}(1) \right) \int_0^1 W_\varepsilon(s) ds. \end{aligned}$$

Note that part I is \mathcal{M} -measurable. Under $\mathbb{P}_{0,0}$ the random variables $W_{b_k}(1)$, $k \in \mathbb{N}$, are independent to W_ε and B_{b_k} , $k \in \mathbb{N}$. Indeed, the independence to W_ε holds by construction and the independence to B_{b_k} is a well-known, and easy to verify, property of Brownian bridges. We thus obtain, since $\mathcal{I}_f(h, \eta)$ is \mathcal{M} -measurable as well,

$$\mathbb{E}_0 \left[\frac{d\mathbb{P}_{h,\eta}}{d\mathbb{P}_{0,\eta}} \mid \mathcal{M} \right] = \exp \left(h \times I - \frac{1}{2} \mathcal{I}_f(h, \eta) \right)$$

$$\begin{aligned}
& \times \mathbb{E}_{0,0} \left[\exp \left(\sum_{k=1}^{\infty} (hJ_{f,k} \int_0^1 W_{\varepsilon}(s) ds + \eta_k) W_{b_k}(1) \right) \mid \mathcal{M} \right] \\
& = \exp \left(h \times I - \frac{1}{2} \mathcal{I}_f(h, \eta) + \frac{1}{2} \sum_{k=1}^{\infty} (hJ_{f,k} \int_0^1 W_{\varepsilon}(s) ds + \eta_k)^2 \right).
\end{aligned}$$

This yields

$$\frac{d\mathbb{P}_h^{\mathcal{M}}}{d\mathbb{P}_0^{\mathcal{M}}} = \exp \left(h\Delta_f^* - \frac{1}{2} h^2 \mathcal{I}_f^* \right)$$

where

$$\Delta_f^* = \int_0^1 W_{\varepsilon}(s) dB_{\phi_f}(s) + W_{\varepsilon}(1) \int_0^1 W_{\varepsilon}(s) ds, \quad (2.3.10)$$

$$\begin{aligned}
\mathcal{I}_f^* &= J_f \int_0^1 W_{\varepsilon}^2(s) ds - \left(\int_0^1 W_{\varepsilon}(s) ds \right)^2 \sum_{k=1}^{\infty} J_{f,k}^2 \\
&= J_f \int_0^1 W_{\varepsilon}^2(s) ds - \left(\int_0^1 W_{\varepsilon}(s) ds \right)^2 (J_f - 1),
\end{aligned} \quad (2.3.11)$$

where the last equality follows from (2.3.4). Note that this likelihood ratio is indeed invariant with respect to η and one can also verify directly that \mathcal{I}_f^* is the quadratic counterpart of Δ_f^* .

We can now formalize the notion of point-optimal invariant tests in the limit experiment. To that end, let us denote the $(1 - \alpha)$ -quantile of $d\mathbb{P}_h^{\mathcal{M}}/d\mathbb{P}_0^{\mathcal{M}}$ under $\mathbb{P}_{0,\eta}$, which does not depend on η , by $c(h, f; \alpha)$. Define the size- α test $\phi_{f,\alpha}^*(\bar{h}) = \mathbb{1} \{ d\mathbb{P}_h^{\mathcal{M}}/d\mathbb{P}_0^{\mathcal{M}} \geq c(\bar{h}, f; \alpha) \}$, for a fixed value of $\bar{h} < 0$. Note that this is an oracle test depending on f , and the feasible version will be provided in Section 2.4. The power function of this oracle test is given by

$$h \mapsto \pi_{f,\alpha}^*(h; \bar{h}) = \mathbb{E}_0 \left[\phi_{f,\alpha}^*(\bar{h}) \frac{d\mathbb{P}_{h,\eta}}{d\mathbb{P}_{0,0}} \right].$$

An application of the Neyman-Pearson lemma yields the following corollary.

Corollary 2.3.2. *Let $f \in \mathcal{F}$ and $\alpha \in (0, 1)$. Let ϕ be a (possibly randomized) test that is \mathcal{M} -measurable and is of size α , i.e., $\mathbb{E}_0 \phi \leq \alpha$. Let π denote the power function of this test, i.e., $\pi(h) = \mathbb{E}_h \phi$. Then we have*

$$\pi(\bar{h}) \leq \pi_{f,\alpha}^*(\bar{h}; \bar{h}).$$

The (oracle) test $\phi_{f,\alpha}^*(\bar{h})$ thus is point optimal, i.e., its power function is tangent to the (semiparametric) power envelope⁵ $h \mapsto \pi_{f,\alpha}^*(h; h)$ at $h = \bar{h}$.

Remark 2.3.3. The semiparametric power envelope derived above, of course, coincides with the one in Jansson (2008) developed based on the invariance constraint. This can be seen by rewriting $\Delta_f^* = \int_0^1 W_\varepsilon(s) dW_{\phi_f}(s) - (W_{\phi_f}(1) - W_\varepsilon(1)) \int_0^1 W_\varepsilon(s) ds$. Our approach is attractive since, by describing the perturbations on f with an orthonormal basis and a infinite-dimensional parameter instead of one single parameter, we don't need to find the least favorable direction. Also, our approach shows how to exploit the invariance constraint, rather than the similarity constraint, in this problem. Furthermore, we show by our approach the possibility for adaptive unit root testing for cases with density known to be symmetric in Remark 2.6.1.

Remark 2.3.4. The semiparametric power envelope $\pi_{f,\alpha}^*$ of the limit experiment in Proposition 2.3.3 is scale invariant, i.e., invariant with respect to the value of $\sigma_f > 0$. This is easily seen from the fact that W_ε , W_{ϕ_f} and J_f are all scale invariant.

Remark 2.3.5. The notion of invariance in the limit experiment leads to another interpretation of the ERS test statistic. Note that σ -field $\mathcal{M}_\varepsilon = \sigma(W_\varepsilon(s); s \in [0, 1])$ is also invariant, though not maximally so. We now calculate the likelihood ratio conditional on observing \mathcal{M}_ε only, by further projecting the likelihood ratio on \mathcal{M}_ε :

$$\frac{d\mathbb{P}_h^{\mathcal{M}_\varepsilon}}{d\mathbb{P}_0^{\mathcal{M}_\varepsilon}} = \mathbb{E}_0 \left[\frac{d\mathbb{P}_h^{\mathcal{M}}}{d\mathbb{P}_0^{\mathcal{M}}} \mid \mathcal{M}_\varepsilon \right]$$

⁵Here and later in this section, the early usage of the concept “power envelope” (instead of “upper bound”) is due to the fact that it is shown to be pointwise attainable in Section 2.4.

$$\begin{aligned}
&= \exp \left(h \int_0^1 W_\varepsilon(s) dB_\varepsilon(s) + h W_\varepsilon(1) \int_0^1 W_\varepsilon(s) ds - \frac{1}{2} h^2 \mathcal{I}_f^* \right) \\
&\quad \times \mathbb{E}_0 \left[\exp \left(h \int_0^1 W_\varepsilon(s) dB_b(s) \right) \mid \mathcal{M}_\varepsilon \right] \\
&= \exp \left(h \int_0^1 W_\varepsilon(s) dW_\varepsilon(s) - \frac{1}{2} h^2 \mathcal{I}_f^* \right) \\
&\quad \times \exp \left(\frac{1}{2} h^2 \left[\int_0^1 W_\varepsilon^2(s) ds - \left(\int_0^1 W_\varepsilon(s) ds \right)^2 \right] (J_f - 1) \right) \\
&= \exp \left(h \int_0^1 W_\varepsilon(s) dW_\varepsilon(s) - \frac{1}{2} h^2 \int_0^1 W_\varepsilon^2(s) ds \right),
\end{aligned}$$

where $W_b(s) = \sum_{k=1}^\infty J_{f,k} W_{b_k}(s)$ and $B_b(s) = B^{W_b}(s)$ for notational simplicity. As a result, the ERS test statistic equals the likelihood ratio statistic from using the (non-maximal) invariant \mathcal{M}_ε . This is an alternative explanation for the improved power of our tests. Moreover, for Gaussian f , we have $\mathcal{M}_\varepsilon = \mathcal{M}$ and obtain point-optimality of the ERS test for this case.⁶

2.3.4 The asymptotic power envelope for asymptotically invariant tests

Now we translate the results for the limiting LABF experiment to the unit root model of interest. To mimick the invariance in the limit experiment we introduce the following definition.

Definition 2.3.1. A sequence of test statistics $\psi^{(T)}$ is said to be asymptotically invariant if the distribution of $\psi^{(T)}$ weakly converges, under $P_{h,\eta;\mu,f}^{(T)}$ for all $h \leq 0$ and $\eta \in c_{00}$, to the distribution of an invariant test in the limit experiment $\mathcal{E}(f)$, under $\mathbb{P}_{h,\eta}$.

The Asymptotic Representation Theorem (see, e.g., Van der Vaart (2000))

⁶Similarly, one could try to derive the statistic resulting from using $\mathcal{M}_B = \sigma(B_{b_k}(s); s \in [0, 1])$ as an invariant. However, that does not seem to lead to an insightful result.

Chapter 9) now yields the following main result on the asymptotic power envelope.

Theorem 2.3.2. *Let $f \in \mathcal{F}$, $\mu \in \mathbb{R}$, and $\alpha \in (0, 1)$. Let $\phi_T(Y_1, \dots, Y_T)$, $T \in \mathbb{N}$, be an asymptotically invariant test of size α , i.e., $\limsup_{T \rightarrow \infty} \mathbb{E}_{0,\eta} \phi_T \leq \alpha$ for all $\eta \in c_{00}$. Let π_T denote the power function of ϕ_T , i.e., $\pi_T(h, \eta) = \mathbb{E}_{h,\eta} \phi_T$. Then we have*

$$\limsup_{T \rightarrow \infty} \pi_T(h, \eta) \leq \pi_{f,\alpha}^*(h; h), \quad \eta \in c_{00} \text{ and } h < 0.$$

The power envelope for invariant tests in the limit experiment thus provides an upper bound to the asymptotic power of invariant tests for the unit root hypothesis. The next section introduces a class of tests that attains this bound (point-wise) and, thereby, demonstrates that the bound indeed constitutes the asymptotic power envelope for invariant unit root tests. We also provide a Chernoff-Savage type result for this class of tests.

2.4 Semiparametrically optimal hybrid rank tests

The appearance of the bridge process B_{ϕ_f} in the “efficient central sequence” Δ_f^* naturally suggests the (partial) use of ranks in the construction of feasible test statistics. Indeed, we can construct an empirical analogue of B_{ϕ_f} by considering a partial-sum process which only depends on the observations via the ranks R_t of ΔY_t amongst $\Delta Y_2, \dots, \Delta Y_T$. We allow for the use of a *reference density* g that may or may not be equal to the true underlying innovation density f . Our findings compare to Quasi-ML methods: if the true innovation density happens to be the same as the selected reference density the inference procedure is point-optimal. At the same time, the procedure is valid, i.e., has proper asymptotic size, even in case the true innovation density does not coincide with the reference density. Note that

these results also hold in case the reference density is non-Gaussian, while Quasi-ML results are generally restricted to Gaussian reference densities.

We need the following mild assumption on the reference density.

Assumption 2.4.1. The density $g \in \mathcal{F}$, with finite variance σ_g^2 , satisfies

$$\lim_{T \rightarrow \infty} \frac{\sigma_g^2}{T} \sum_{i=1}^T \phi_g^2 \left(G^{-1} \left(\frac{i}{T+1} \right) \right) = J_g,$$

with location score function $\phi_g(e) := -(g'/g)(e)$, where J_g is the standardized Fisher information⁷ for location of g .

Now we can formulate the following direct extension of Lemma A.1 in Hallin, Van den Akker, and Werker (2011). The proof is omitted.

Lemma 2.4.1. *Let $f \in \mathcal{F}$, $\mu \in \mathbb{R}$, and g satisfy Assumption 2.4.1. Consider the partial sum process*

$$B_{\phi_g}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \sigma_g \phi_g \left(G^{-1} \left(\frac{R_t}{T+1} \right) \right), \quad s \in [0, 1], \quad (2.4.1)$$

where R_t denotes the rank of ΔY_t , $t = 2, \dots, T$. Then, under $P_{0,0;\mu,f}^{(T)}$ and as $T \rightarrow \infty$, we have⁸

$$\begin{bmatrix} W_{\varepsilon}^{(T)} \\ W_{\phi_f}^{(T)} \\ B_{\phi_g}^{(T)} \end{bmatrix} \Rightarrow \begin{bmatrix} W_{\varepsilon} \\ W_{\phi_f} \\ B_{\phi_g} \end{bmatrix} \quad \text{and} \quad (2.4.2)$$

$$\int_0^1 W_{\varepsilon}^{(T)}(s-) dB_{\phi_g}^{(T)}(s) \Rightarrow \int_0^1 W_{\varepsilon}(s) dB_{\phi_g}(s). \quad (2.4.3)$$

Here, B_{ϕ_g} is the associated Brownian bridge of W_{ϕ_g} , which itself is a Brownian motion defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0})$ as W_{ε} and

⁷Similarly as the standardized Fisher information J_f of f , Fisher information J_g of g is also standardized, by the variance σ_g^2 , so that it is scale invariant.

⁸Equation (2.4.3) holds because of the Theorem 2.1 in Hansen (1992).

W_{ϕ_f} , with covariance matrix

$$\text{Cov}_{0,0} \begin{pmatrix} W_\varepsilon(1) \\ W_{\phi_f}(1) \\ W_{\phi_g}(1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sigma_{\varepsilon\phi_g} \\ & J_f & J_{fg} \\ & & J_g \end{pmatrix}, \quad (2.4.4)$$

where

$$\begin{aligned} \sigma_{\varepsilon\phi_g} &= \sigma_f^{-1} \sigma_g \int_0^1 F^{-1}(u) \phi_g(G^{-1}(u)) du, \\ J_{fg} &= \sigma_f \sigma_g \int_0^1 \phi_f(F^{-1}(u)) \phi_g(G^{-1}(u)) du. \end{aligned}$$

2.4.1 The hybrid rank tests based on a reference density

The weak convergence in Lemma 2.4.1 indicates that constructing the partial sum processes as described above (with rank statistics and a reference density g for $B_{\phi_g}^{(T)}$), as T approaches infinity, corresponds to observing the σ -field $\mathcal{M}_g = \sigma(W_\varepsilon(s), B_{\phi_g}(s); s \in [0, 1])$ in the limit. Clearly, $\mathcal{M}_g \subseteq \mathcal{M}^9$ so that \mathcal{M}_g is invariant for the group of transformations g_η . When $g = f$, $\mathcal{M}_g = \mathcal{M}$ so that it is maximally invariant, which means that we capture all available information about h .

The following proposition establishes the likelihood ratio restricted to the information \mathcal{M}_g .

Proposition 2.4.1. *Define W_\perp via the decomposition*

$$\frac{W_{\phi_g}}{\sigma_{\varepsilon\phi_g}} = W_\varepsilon + \sqrt{\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1} W_\perp, \quad (2.4.5)$$

which is the standard Brownian motion under $\mathbb{P}_{0,0}$, and denote the associated Brownian bridge by B_\perp . The likelihood ratio $d\mathbb{P}_h/d\mathbb{P}_0$ restricted to the

⁹This is due to the decomposition $B_{\phi_g} = \sigma_{\varepsilon\phi_g} B_\varepsilon + \sum_{k=1}^\infty J_{g,k} B_{b_k}$.

outcome space \mathcal{M}_g is given by

$$\frac{d\mathbb{P}_h^{\mathcal{M}_g}}{d\mathbb{P}_0^{\mathcal{M}_g}} = \mathbb{E}_0 \left[\frac{d\mathbb{P}_h^{\mathcal{M}}}{d\mathbb{P}_0^{\mathcal{M}}} \mid \mathcal{M}_g \right] = \exp \left(h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g \right), \quad (2.4.6)$$

with

$$\Delta_g = \Delta_\varepsilon + \lambda\Delta_\perp,$$

$$\mathcal{I}_g = \int_0^1 W_\varepsilon^2(s)ds + \lambda^2 \left(\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1 \right) \left[\int_0^1 W_\varepsilon(s)^2ds - \left(\int_0^1 W_\varepsilon(s)ds \right)^2 \right],$$

where $\Delta_\varepsilon = \int_0^1 W_\varepsilon(s)dW_\varepsilon(s)$, $\Delta_\perp = \sqrt{J_g/\sigma_{\varepsilon\phi_g}^2 - 1} \int_0^1 W_\varepsilon(s)dB_\perp(s)$ and $\lambda = (J_{fg}\sigma_{\varepsilon\phi_g} - \sigma_{\varepsilon\phi_g}^2)/(J_g - \sigma_{\varepsilon\phi_g}^2)$.

Remark 2.4.1. The result of Proposition 2.4.1 can also be achieved by firstly applying Girsanov's Theorem to the following experiment

$$\begin{aligned} dW_\varepsilon(s) &= hW_\varepsilon(s)ds + dZ_\varepsilon(s), \\ dW_{\phi_g}(s) &= hJ_{fg}W_\varepsilon(s)ds + \eta_g ds + dZ_{\phi_g}(s), \end{aligned}$$

to get the likelihood ratio of $\sigma(W_\varepsilon(s), W_{\phi_g}(s), s \in [0, 1])$ and, subsequently, taking the expectation of it conditional on \mathcal{M}_g . The experiment above is obtained by combining the limit experiment in Proposition 2.3.3 and the covariance matrix in (4.2.15). Here $\eta_g = \sum_k \eta_k J_{g,k}$ with $J_{g,k} = \text{Cov}_{0,0}(W_{\phi_g}(1), W_{b_k}(1))$.

Observe that W_\perp is a standard Brownian motion under $P_{0,0;\mu,f}^{(T)}$ which is independent of W_ε . When $g = f$, we have $J_{fg} = J_f = J_g$ and $\sigma_{\varepsilon\phi_g} = 1$, so that $\lambda = 1$ and $B_{\phi_g} = B_{\phi_f}$. As a result $\Delta_g = \Delta_f^*$ and $\mathcal{I}_g = \mathcal{I}_f^*$.

The central idea to construct a hybrid rank test is to use a (quasi)-log-likelihood ratio test based on $L_{\mathcal{M}_g}(h, \lambda) = h\Delta_g - \frac{1}{2}h^2\mathcal{I}_g$ from (2.4.6), where we replace W_ε and B_{ϕ_g} by their finite-sample counterparts from Lemma 2.4.1 and the unknown parameters σ_f^2 and λ by estimates. Therefore, we impose the following condition.

Assumption 2.4.2. There exist consistent, under the null hypothesis, estimators $\hat{\sigma}_f^2 > 0$ a.s., $\hat{\sigma}_{\varepsilon\phi_g}$, and \hat{J}_{fg} of σ_f^2 , $\sigma_{\varepsilon\phi_g}$, and J_{fg} , respectively. More precisely, for all $f \in \mathcal{F}$, we have $\hat{\sigma}_f^2 \xrightarrow{P} \sigma_f^2$, $\hat{\sigma}_{\varepsilon\phi_g} \xrightarrow{P} \sigma_{\varepsilon\phi_g}$, and $\hat{J}_{fg} \xrightarrow{P} J_{fg}$, under $P_{0,0;\mu,f}^{(T)}$ as $T \rightarrow \infty$.

Such estimators are easily constructed, although \hat{J}_{fg} is somewhat more involved. Estimating the *real-valued* cross-information J_{fg} requires non-parametric techniques, but is considerably simpler than a full *nonparametric* estimation of ϕ_f . Estimating J_{fg} can be done along similar lines as estimating the Fisher information J_f , see, e.g., Bickel (1982), Bickel et al. (1998), Schick (1986), and Klaassen (1987). A direct rank-based estimator of J_{fg} has been proposed in Cassart, Hallin, Paindaveine (2010).

Next, based on a chosen reference density g satisfying Assumption 2.4.1 and some estimators $\hat{\sigma}_f$, $\hat{\sigma}_{\varepsilon\phi_g}$ and \hat{J}_{fg} satisfying Assumption 2.4.2, we introduce the following two partial sum processes:

$$\begin{aligned}\widehat{W}_\varepsilon^{(T)}(s) &= \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \frac{\Delta Y_t}{\hat{\sigma}_f}, \\ \widehat{B}_\perp^{(T)}(s) &= \left(\frac{J_g}{\hat{\sigma}_{\varepsilon\phi_g}^2} - 1 \right)^{-\frac{1}{2}} \left[\frac{B_{\phi_g}^{(T)}(s)}{\hat{\sigma}_{\varepsilon\phi_g}} - \left(\widehat{W}_\varepsilon^{(T)}(s) - \widehat{W}_\varepsilon^{(T)}(1)\lfloor sT \rfloor \right) \right],\end{aligned}$$

where $B_{\phi_g}^{(T)}(s)$ is defined in (2.4.1). Then, given a fixed alternative $\bar{h} < 0$, we define

$$\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) := \bar{h} \widehat{\Delta}_g^{(T)} - \frac{1}{2} \bar{h}^2 \widehat{\mathcal{I}}_g^{(T)}, \quad (2.4.7)$$

with

$$\begin{aligned}\widehat{\Delta}_g^{(T)} &= \widehat{\Delta}_\varepsilon^{(T)} + \hat{\lambda} \widehat{\Delta}_\perp^{(T)}, \\ \widehat{\mathcal{I}}_g^{(T)} &= \int_0^1 \left(\widehat{W}_\varepsilon^{(T)}(s-) \right)^2 ds \\ &\quad + \hat{\lambda}^2 \left(\frac{J_g}{\hat{\sigma}_{\varepsilon\phi_g}^2} - 1 \right) \left[\int_0^1 \left(\widehat{W}_\varepsilon^{(T)}(s-) \right)^2 ds - \left(\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) ds \right)^2 \right],\end{aligned}$$

where

$$\begin{aligned}\widehat{\Delta}_\varepsilon^{(T)} &= \int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) d\widehat{W}_\varepsilon^{(T)}(s), \\ \widehat{\Delta}_\perp^{(T)} &= \sqrt{J_g/\hat{\sigma}_{\varepsilon\phi_g}^2 - 1} \int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) d\widehat{B}_\perp^{(T)}(s), \\ \hat{\lambda} &= (\hat{J}_{fg}\hat{\sigma}_{\varepsilon\phi_g} - \hat{\sigma}_{\varepsilon\phi_g}^2)/(J_g - \hat{\sigma}_{\varepsilon\phi_g}^2).\end{aligned}$$

By Slutsky's theorem and (2.4.2), we have $(\widehat{W}_\varepsilon^{(T)}, \widehat{B}_{\phi_g}^{(T)})' \Rightarrow (W_\varepsilon, B_{\phi_g})'$, and thus $(\widehat{W}_\varepsilon^{(T)}, \widehat{B}_\perp^{(T)})' \Rightarrow (W_\varepsilon, B_\perp)'$. It follows $\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) \Rightarrow L_{\mathcal{M}_g}(\bar{h}, \lambda)$. Define the critical value $c_{\mathcal{M}_g}(\bar{h}, \sigma_{\varepsilon\phi_g}, \lambda, J_g; \alpha)$ by the $(1 - \alpha)$ -quantile of $L_{\mathcal{M}_g}(\bar{h}, \lambda)$. This leads to the feasible test

$$\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha) := \mathbb{1} \left\{ \widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda}) \geq c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha) \right\}.$$

Since these tests are based on the ranks of ΔY_t , but also their average, we name them Hybrid Rank Tests (HRTs). We can now state our main theoretical result.

Theorem 2.4.1. *Choose $\alpha \in (0, 1)$, $\bar{h} \in (-\infty, 0)$, and g satisfying Assumption 2.4.1. For each $\mu \in \mathbb{R}$, $h \in (-\infty, 0)$ and $f \in \mathcal{F}$, we have:*

- (i) *The Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$ is asymptotically of size α .*
- (ii) *The Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$ is asymptotically invariant.*
- (iii) *The Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$ is point-optimal, at $h = \bar{h}$, if $g = f$.*

Theorem 2.4.1 shows the HRTs are valid irrespective of the choice of the reference density and point-optimal for a correctly specified reference density. The proof of this theorem is based on weak convergence of the test statistic $\widehat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda})$ to its limit $L_{\mathcal{M}_g}(\bar{h}, \lambda)$ as shown above. Then, (i) is derived directly by the design of the test; (ii) is proved by the fact that $L_{\mathcal{M}_g}(\bar{h}, \lambda)$ is \mathcal{M} -measurable. The proof of (iii) comes from last part of Proposition 2.4.1: when $g = f$, $L_{\mathcal{M}_g}(h, \lambda) = \log(d\mathbb{P}_h^{\mathcal{M}}/d\mathbb{P}_0^{\mathcal{M}})$.

Corollary 2.4.1 (Chernoff-Savage type result). *Fix $\alpha \in (0, 1)$ and $\bar{h} < 0$. The Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$ is, for any reference density g satisfying Assumption 2.4.1, more powerful, at $h = \bar{h}$ and for $\mu \in \mathbb{R}$ and $f \in \mathcal{F}$, than the ERS test. Both tests have equal power if f is Gaussian.*

Proof. Recall once more that $\hat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, \hat{\lambda})$ weakly converges to $L_{\mathcal{M}_g}(\bar{h}, \lambda)$, which, in the limit, is the likelihood ratio restricted to the σ -field \mathcal{M}_g . Then, by the Neyman-Pearson Lemma, we conclude from $\mathcal{M}_\varepsilon \subseteq \mathcal{M}_g$ that the HRT is more powerful than the ERS test at $h = \bar{h}$. Recalling the decomposition (3.2.9), write the limit experiment in Remark 2.4.1 as

$$\begin{aligned} dW_\varepsilon(s) &= hW_\varepsilon(s)ds + dZ_\varepsilon(s), \\ dW_\perp(s) &= h \frac{J_{fg} - \sigma_{\varepsilon\phi_g}}{\sqrt{J_g - \sigma_{\varepsilon\phi_g}^2}} W_\varepsilon(s)ds + \frac{\eta_g}{\sqrt{J_g - \sigma_{\varepsilon\phi_g}^2}} ds + dZ_\perp(s). \end{aligned}$$

When f is Gaussian, W_\perp (or W_{ϕ_g}) provides no more information about h than W_ε since $J_{fg} = \sigma_{\varepsilon\phi_g}$. In that case the HRT and the ERS test are asymptotically equivalent. \square

Corollary 2.4.1 is a particularly useful result for applied work. The HRT dominates its classical canonical Gaussian counterpart, i.e., the ERS test in the present model, for any reference density g . Traditionally, this claim can only be made for Gaussian reference densities, but the non-LAN framework here even allows for a stronger result. Our formulation of the testing problem using invariance arguments is convenient in this respect: the larger the invariant σ -field that is used, the more powerful the test.

The situation can be compared to Quasi Maximum Likelihood methods. However, again, in classical situations these methods are generally restricted to Gaussian reference densities. In the present setup, any reference density g (subject to the regularity conditions imposed) can be used. The resulting test will always be valid, but more powerful in case the reference density chosen is closer to the true underlying density f .

Remark 2.4.2. The additional power of the HRT compared to the ERS test is not free of charge due to the stronger weak convergence assumption employed. Consequently, the class of models for which the HRTs are valid forms a sub-class of the class where the ERS tests are valid. In this sub-class, the HRT dominates the ERS test, but outside they may even lose validity. In the opposite direction, the Müller and Watson (2008) low-frequency unit root test can be applied in a even larger class of models than the ERS tests. Again, within the class of models where the ERS test is valid, it has lower power. A more general and detailed discussion in this direction can be found in Müller (2011).

Our test will still be relevant in many applications, notably those where policy implications are derived under an i.i.d. assumption on the innovations. Also, our approach can most likely be extended to situations where the innovations are described by some explicit dynamic location-scale model. We come back to this point in Section 3.6.

2.4.2 The approximate hybrid rank tests

A somewhat inconvenient aspect of the hybrid rank tests is that we need to estimate J_{fg} . As mentioned before, this is (much) less complicated than estimating the score function ϕ_f , but might still be considered cumbersome, despite all references mentioned below Assumption 2.4.2. Moreover, the critical value $c_{\mathcal{M}_g}(\bar{h}, \hat{\sigma}_{\varepsilon\phi_g}, \hat{\lambda}, J_g; \alpha)$ depends on estimates $\hat{\sigma}_{\varepsilon\phi_g}$ and $\hat{\lambda}$ (henceforth \hat{J}_{fg}). This introduces no difficulty to implementing the test, however, when it comes to simulations, the computational effort will be significant. Therefore, we introduce additionally a simplified version of the hybrid rank test. This simplified test is obtained by setting $\lambda = 1$, which holds in case $g = f$.

To be precise, define

$$\hat{L}_g^{(T)}(\bar{h}) := \hat{L}_{\mathcal{M}_g}^{(T)}(\bar{h}, 1) = \bar{h} \hat{\Delta}_g^{(T)} - \frac{1}{2} \bar{h}^2 \hat{\mathcal{I}}_g^{(T)}, \quad (2.4.8)$$

where

$$\begin{aligned}\widehat{\Delta}_g^{(T)} &= \frac{1}{\widehat{\sigma}_{\varepsilon\phi_g}} \int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) d\widehat{B}_{\phi_g}^{(T)}(s) + \widehat{W}_\varepsilon^{(T)}(1) \int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) ds, \\ \widehat{\mathcal{I}}_g^{(T)} &= \frac{J_g}{\widehat{\sigma}_{\varepsilon\phi_g}^2} \int_0^1 \left(\widehat{W}_\varepsilon^{(T)}(s-) \right)^2 ds - \left(\int_0^1 \widehat{W}_\varepsilon^{(T)}(s-) ds \right)^2 \left(\frac{J_g}{\widehat{\sigma}_{\varepsilon\phi_g}^2} - 1 \right),\end{aligned}$$

and $L_g(\bar{h}) := L_{\mathcal{M}_g}(\bar{h}, 1)$. By the same arguments as before, we have $\widehat{L}_g^{(T)}(\bar{h}) \Rightarrow L_g(\bar{h})$. Denoting the $(1 - \alpha)$ -quantile of $L_g(\bar{h})$ by $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$, this leads to the feasible test

$$\phi_g^{(T)}(\bar{h}, \alpha) := \mathbb{1} \left\{ L_g^{(T)}(\bar{h}) \geq c_g(\bar{h}, \widehat{\sigma}_{\varepsilon\phi_g}, J_g; \alpha) \right\}.$$

Since $\phi_g^{(T)}(\bar{h}, \alpha)$ is an approximate version of the Hybrid Rank Test $\phi_{\mathcal{M}_g}^{(T)}(\bar{h}, \alpha)$, we refer to it as Approximate Hybrid Rank Test (AHRT).

Theorem 2.4.2. *Under the same conditions as Theorem 2.4.1, the asymptotic properties of the Hybrid Rank Tests — validity, invariance, and point-optimality when $g = f$ — also hold for the Approximate Hybrid Rank Tests.*

The proof of Theorem 2.4.2 follows along the same lines as that of Theorem 2.4.1 but using the weak convergence $\widehat{L}_g^{(T)}(\bar{h}) \Rightarrow L_g(\bar{h})$. The simulation results in Section 4.5 show that these asymptotic properties carry over to finite samples.

From a computational point of view, the AHRT has the advantage that nonparametric estimation of J_{fg} is no longer needed. This significantly reduces the computational effort in the Monte-Carlo study. Indeed, even though the critical value $c_g(\bar{h}, \sigma_{\varepsilon\phi_g}, J_g; \alpha)$ is still data dependent, it is, for given α , \bar{h} , and reference density g , a function of only one argument — the parameter $\sigma_{\varepsilon\phi_g}$. Observe, by Cauchy-Schwarz, that $\sigma_{\varepsilon\phi_g}$ is bounded by $\sqrt{J_g}$. For the chosen three reference densities, the critical value functions are listed in Table 2.1. In the Section 4.5, we use these estimated critical value functions for computational speed.

Table 2.1: This table provides estimated critical value functions for three reference densities: Gaussian ($J_g = 1$), Laplace ($J_g = 2$), and Student t_3 ($J_g = 2$) at $\alpha = 5\%$ and $\bar{h} = -7\sigma_{\varepsilon\phi_g}$. For each case, the critical value function is estimated by OLS regression using simulated critical values on the interval $[0, \sqrt{J_g}]$ with a grid where adjacent points are 0.01 apart.

g	
Gaussian	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 1; 5\%) = 0.96 + 1.88\sigma_{\varepsilon\phi_g} - 3.98\sigma_{\varepsilon\phi_g}^2 + 6.74\sigma_{\varepsilon\phi_g}^3 - 5.45\sigma_{\varepsilon\phi_g}^4$
Laplace	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 2; 5\%) = 0.25 + 2.30\sigma_{\varepsilon\phi_g} - 3.58\sigma_{\varepsilon\phi_g}^2 + 4.30\sigma_{\varepsilon\phi_g}^3 - 2.45\sigma_{\varepsilon\phi_g}^4$
t_3	$c_g(-7\sigma_{\varepsilon\phi_g}, \sigma_{\varepsilon\phi_g}, 2; 5\%) = 0.25 + 2.30\sigma_{\varepsilon\phi_g} - 3.58\sigma_{\varepsilon\phi_g}^2 + 4.30\sigma_{\varepsilon\phi_g}^3 - 2.45\sigma_{\varepsilon\phi_g}^4$

Remark 2.4.3 (Chernoff-Savage result for the AHRTs). Although we are not able to provide a rigorous mathematical proof, the Monte-Carlo study indicates that the Chernoff-Savage property is also preserved for the AHRT, at least in case the reference density g is chosen to be Gaussian. Such a result would be more in line with applications of the Chernoff-Savage result in classical LAN situations.

Remark 2.4.4. It is worth noting that the invariance constraint is only imposed in the limit and, henceforth the maximal invariant needs only to be derived in the limit experiment. In other words, in the finite-sample unit root testing experiment $\mathcal{E}^{(T)}(f)$, we actually use statistics that are only asymptotically invariant (i.e., their limiting equivalents are measurable with respect to the maximally invariant sigma-field $\mathcal{M} = \sigma(W_\varepsilon, (B_{b_k})_{k \in \mathbb{N}})$), while, for finite T , they are not necessarily (maximally) invariant w.r.t. some transformation on the density f . In fact, such maximal invariant may very well not even exist the finite-sample experiments. Specifically, $W_\varepsilon^{(T)}$ approximates W_ε in the limit, whose distribution does not change with the density f , while the distribution of $W_\varepsilon^{(T)}$ does depend on the density f . Instead, the statistic $B_{\phi_g}^{(T)}$ is distribution-free, that is, its distribution is not affected by any transformation on the density f . In Section 4.5 below, we will show that the asymptotic approximations work well even in smaller samples.

2.5 Monte Carlo study

This section reports the results of a Monte Carlo study to corroborate our asymptotic results, and to analyze the small-sample performances of the Approximate Hybrid Rank Tests. As mentioned earlier, we use the Approximate Hybrid Rank Tests in this simulation to avoid having to simulate the critical value for each individual replication. For the fixed alternative, we choose $\bar{h} = -7\sigma_{\varepsilon\phi_g}$ for two reasons. First, when $g = f$, we have $\sigma_{\varepsilon\phi_g} = 1$ and hence $\bar{h} = -7$, which is in line with Elliott, Rothenberg, and Stock (1996). Second, if we choose $\bar{h} = -7$, the critical value will approach $-\infty$ when $\sigma_{\varepsilon\phi_g} \rightarrow 0$. The corresponding critical value functions for various reference densities are provided in Table 2.1. The estimators for σ_f^2 and $\sigma_{\varepsilon\phi_g}$ we use are

$$\hat{\sigma}_f^2 = \frac{1}{T-1} \sum_{t=2}^T (\Delta Y_t - \frac{1}{T} \sum_{t=2}^T \Delta Y_t)^2, \quad (2.5.1)$$

$$\hat{\sigma}_{\varepsilon\phi_g} = \frac{1}{T-1} \sum_{t=2}^T \frac{\Delta Y_t}{\hat{\sigma}_f} \sigma_g \phi_g \left(G^{-1} \left(\frac{R_t}{T+1} \right) \right). \quad (2.5.2)$$

Moreover, to simplify the notations, we denote the Approximate Hybrid Rank test with reference density g by AHRT^g and, in particular, by AHRT^ϕ for Gaussian reference density. Throughout we use the significance level $\alpha = 5\%$ and all simulations are based on 20,000 Monte-Carlo repetitions.

We compare our AHRT with two alternatives. First we consider the Dickey-Fuller test (denoted by $\text{DF-}\rho$) from Dickey and Fuller (1979). This test is based on the statistic $T(\hat{\rho} - 1)$ where $\hat{\rho}$ is the least-squares estimator in the regression $Y_t = \mu + \rho Y_{t-1} + \varepsilon_t$. The critical values for this test are -13.52 for $T = 100$ and -14.05 for $T = 2500$. The second competitor is the ERS test with $\bar{h} = -7$. This test is based on the statistic $[S(\bar{\alpha}) - \bar{\alpha}S(1)]/\hat{\omega}^2$ with $\bar{\alpha} = 1 + T^{-1}\bar{h}$ and $S(a) = (Y_a - Z_a\hat{\beta})'(Y_a - Z_a\hat{\beta})$, with Y_a and Z_a

defined as

$$Y_a = (Y_1, Y_2 - aY_1, \dots, Y_T - aY_{T-1})',$$

$$Z_a = (1, 1 - a, \dots, 1 - a)',$$

where $\hat{\beta}$ is estimated by regressing $Y_{\hat{\alpha}}$ on $Z_{\hat{\alpha}}$. Since in the present model we employ the i.i.d. assumption on the innovations, the long-run variance estimator $\hat{\omega}^2$ is chosen to be $\hat{e}'\hat{e}/T$, where \hat{e} is the residual vector from the regression $\Delta Y_t = \mu + \delta Y_{t-1} + \varepsilon_t$. The critical values for this test are 3.11 for $T = 100$ and 3.26 for $T = 2500$. We do not consider the Dickey-Fuller t -test as it is dominated by the DF- ρ test in the current model. Similarly, the DF-GLS test proposed in Elliott, Rothenberg, and Stock (1996) is also omitted as it behaves asymptotically the same as the ERS test, but can be oversized in small samples.

2.5.1 Large-sample performance

In this section, we use the large-sample performances of the three tests mentioned above to illustrate the asymptotic properties. In particular, the chosen sample size T is 2,500.

Figure 2.1 shows the power curves for 9 combinations of 3 innovation densities f and 3 reference densities g . Each are chosen to be Laplace, Student t_3 , or Gaussian. In line with our theoretical results, we find that the AHRT outperforms the two competitors in most cases. More specifically, when $g = f$ (the graphs on the diagonal), the AHRT^f has power very close to the semiparametric power envelope and it is tangent to it at the point $h = -7$. Moreover, when the reference density g is Gaussian (the three right-most graphs), the AHRT^ϕ outperforms the competitors for all three true densities f . This corroborates the Chernoff-Savage property of the AHRT^ϕ test mentioned in Remark 2.4.3. When both g and f are Gaussian, the AHRT test and the ERS test have indistinguishable power.

In order to investigate the Chernoff-Savage result even further, we consider in Figure 2.2 the AHRT^ϕ test for nine true innovation densities f .

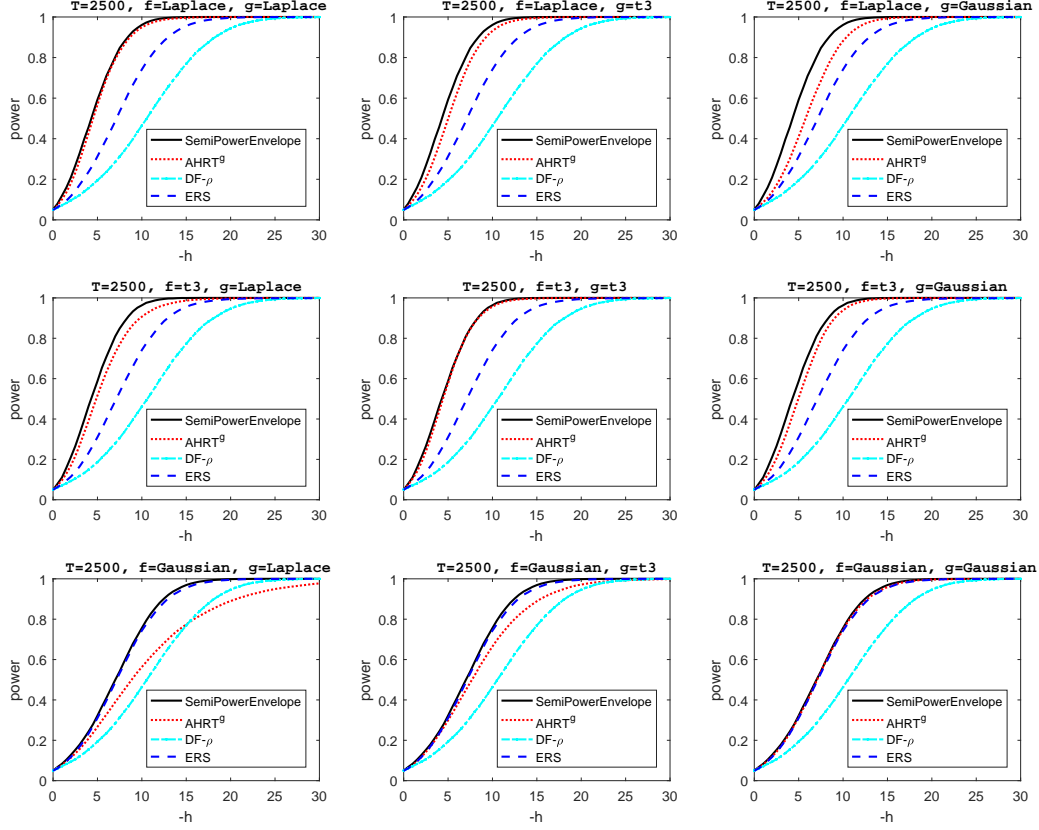


Figure 2.1: Asymptotic power functions of the Hybrid Rank Test for various reference densities g and other selected unit root tests under the true innovation densities f : Gaussian, Laplace, Student's t_3 .

These include innovation densities f that are extremely heavy-tailed, skewed, or both. The first row of graphs shows three extremely heavy-tailed distributions: Student t_2 , Student t_1 , and a stable distribution with stability parameter $\alpha = 0.5$, skewness parameter $\beta = 0$, scale parameter $c = 1$, and location parameter $\mu = 0$. As these densities do not all satisfy our maintained assumptions, these graphs do not show power envelopes.

The top three graphs in Figure 2.2 show that the AHRT^ϕ is much more powerful than its competitors and that its power increases with the heav-

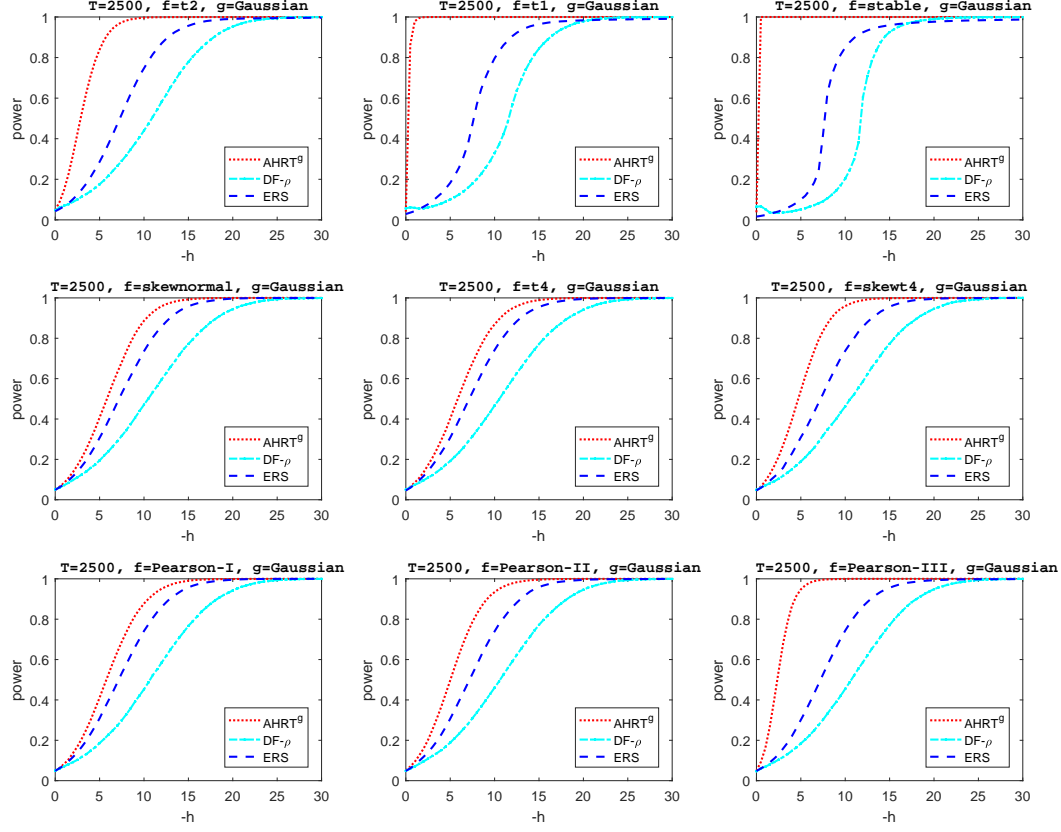


Figure 2.2: Illustration of the Chernoff-Savage result. The figure shows asymptotic power functions for the Gaussian Hybrid Rank Test and selected unit root tests under various true innovation densities f .

iness of the tail. The second and third row show the effect of skewness in f . Specifically, the AHRT^ϕ power function when f is skewed-normal (with skewness 0.8145) is higher than that when f is normal (in Figure 2.1). This indicates that the AHRT^ϕ can acquire power from skewness. The same conclusion can be drawn from the comparison of the AHRT^ϕ power function for t_4 and that of a skewed t_4 with skewness ≈ 2.7 . To further remove the effects of the other moments, in the third row, we also employ the Pearson distributions with identical mean, variance and kurtosis, but different

skewness — skewness = 1 for Pearson-I, skewness = 3 for Pearson-II, and skewness = 6 for Pearson-III. Comparing the corresponding three AHRT^ϕ power functions, it seems that the larger the skewness of the true distribution f is, the more powerful the AHRT^ϕ becomes.

A final remark on the size of the AHRT. In all cases where the true density f satisfies our maintained assumption, i.e., $f \in \mathcal{F}$ (that is all cases in Figure 2.1 and the skewnormal, t_4 , Pearson-I, Pearson-II, and Pearson-III in Figure 2.2), the simulated sizes are between 4.9% and 5.1%. This verifies the validity of the AHRTs claimed in Theorem 2.4.2. In the other cases, i.e., $f \notin \mathcal{F}$, the AHRT is somewhat conservative. More precisely, the simulated sizes of the AHRT^ϕ are 4.845%, 4.085%, 3.725%, and 4.735% for the t_2 , t_1 , stable, and skew- t_4 distribution, respectively. This result seems consistent over all simulations, but we have not been able to provide formal proof.

2.5.2 Small-sample performance

In this section, we report the performance of the AHRTs and the two competitors described above for smaller samples. Figures 2.3 and 2.4 are the small-sample versions, with $T = 100$, of Figures 2.1 and 2.2, respectively. We observe that, even with a slight downward shift of the power functions for all three tests considered, the findings of the large-sample case remain valid in this small-sample case. For larger values of h , the DF- ρ test sometimes dominates the other two tests. This is due to the fact that the DF- ρ in this Monte-Carlo setting has a superior convergence speed (towards its asymptotic power as sample size T increases) to those of the AHRTs and the ERS test, which makes it have a better performance in (ultra)-small-sample cases. In cases with enough samples (say, $T \geq 100$) and when f is significantly away from the Gaussian density, irrespective of the choice of g , the AHRTs performs favorably.

Concerning the small-sample size, we find it to range from about 4.0% to 4.5% for the cases where $f \in \mathcal{F}$. Again, when f does not satisfy our maintained assumptions ($f \notin \mathcal{F}$) the AHRT turns out to be conservative.

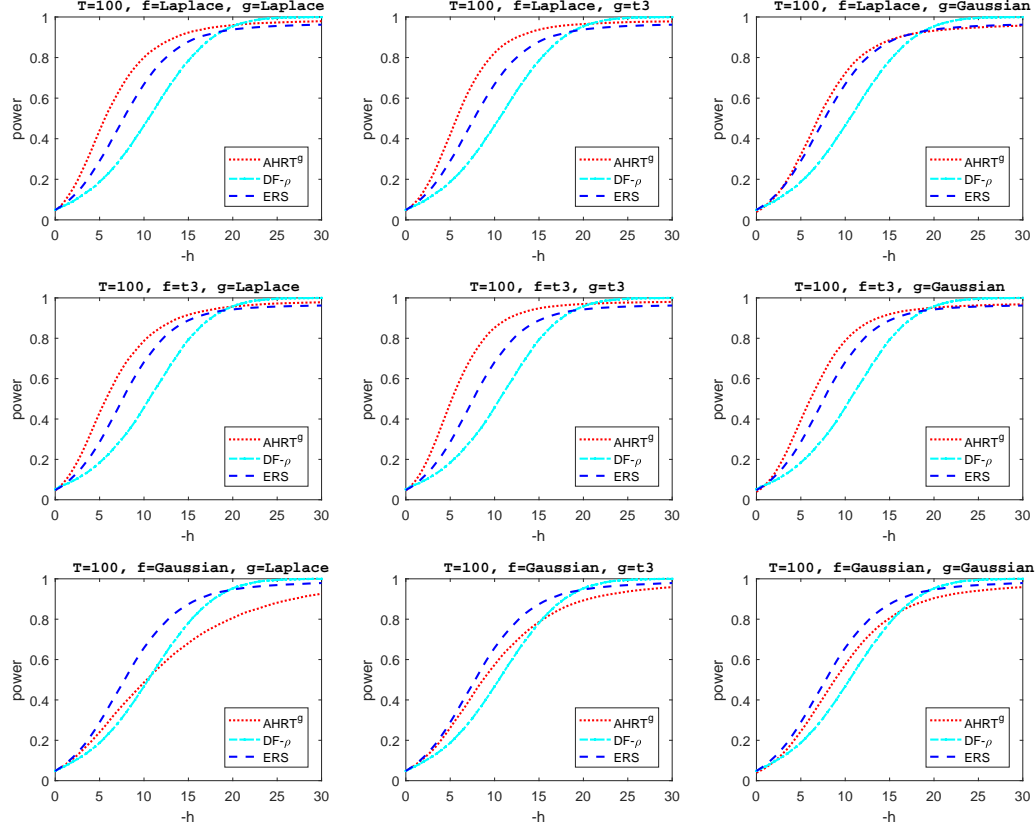


Figure 2.3: Small-sample ($T = 100$) power functions of selected unit root tests and various true innovation densities: Gaussian, Laplace, Student t_3 .

More precisely, we find a size of 3.7%, 3.1%, 2.4%, and 4.1% for the t_2 , t_1 , stable, and skew- t_4 distribution, respectively. This makes the improved power even more remarkable.

It may also be useful to illustrate the convergence of the power function of the AHRT^f to the semiparametric power envelope as sample size T increases. This is the purpose of Figure 2.5. For three cases: Gaussian, Laplace, and Student t_3 , we find that the convergence indeed occurs already at relatively small samples, which is not always the case for alternative unit root tests.

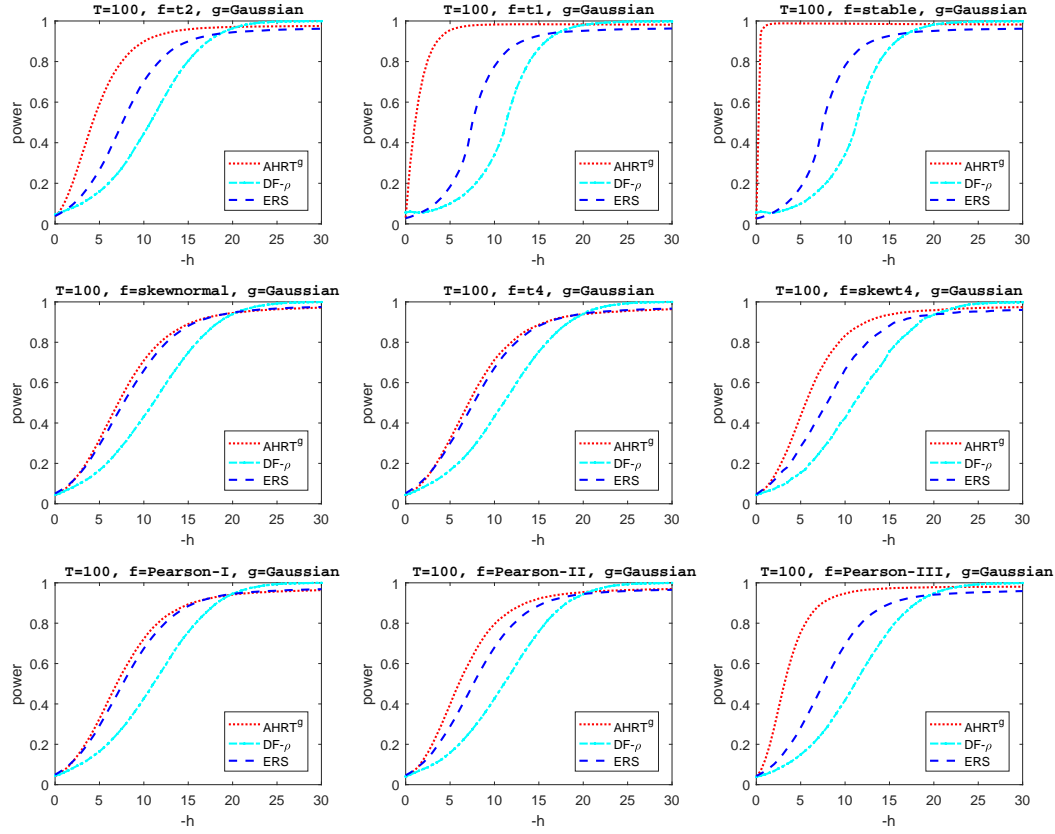


Figure 2.4: Small-sample ($T = 100$) power functions of selected unit root tests and various true innovation densities.

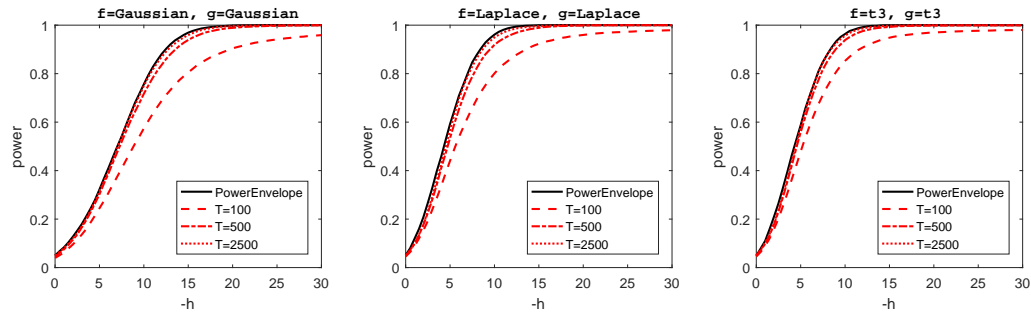


Figure 2.5: Powers of HRT^g when $g = f$ with different sample sizes.

2.6 Discussions of possible extensions

In this section, we discuss three possible extensions of HRTs and AHRTs: a version based on signed rank statistics in case the error distribution is known to be symmetric, a version based on aligned ranks for the case where the innovations ε_t are serially correlated, and a version based on a nonparametrically estimated reference density which addresses globally optimality in \mathcal{F} . Note that these cases may be concurrent and we can combine these extensions accordingly. Detailed proofs are omitted, nevertheless, we provide some relevant Monte-Carlo results of the modified AHRTs in Section 2.6.1.

Remark 2.6.1 (Symmetric error distributions). At first, we consider the case that $f \in \mathcal{F}$ is known to be symmetric. The density f is modeled nonparametrically as in equation (4.2.8), but with all the perturbation scores $b_k(e)$ (could be) chosen as even functions. We have $J_{f,k} = \sigma_f E_f[b_k(\varepsilon)\phi_f(\varepsilon)] = 0$ since the score function ϕ_f is odd. As a consequence, Δ_f is independent of Δ_{b_k} for all k . This gives adaptivity¹⁰ for testing the parameter of interest, h , in the presence of the nuisance parameter η : applying Girsanov's Theorem gives the limit experiment structurally as in Proposition 2.3.3

$$\begin{aligned} dZ_\varepsilon(s) &= dW_\varepsilon(s) - hW_\varepsilon(s)ds, \\ dZ_{\phi_f}(s) &= dW_{\phi_f}(s) - hJ_f W_\varepsilon(s)ds, \\ dZ_{b_k}(s) &= dW_{b_k}(s) - \eta_k ds, \quad k \in \mathbb{N}. \end{aligned}$$

The equation for W_{ϕ_f} is not omitted since (2.3.2) does not hold anymore. By the same method used to prove Theorem 2.3.1, we can show that $\sigma(W_\varepsilon, W_{\phi_f}, B_{b_k})$ is maximal invariant. Subsequently, after some simple algebra, we find the semiparametric power envelope based on this maximal invariant coincides with parametric power envelope where f is known. This verifies again the adaptation result from Jansson (2008) under the same condition by using the new approach. To demonstrate that the semiparametric power envelope

¹⁰A discussion about definition of “adaptive” in this nonstandard unit root testing problem can be found in Section 5 of Jansson (2008).

is sharp, we propose a test based on signed-rank statistics. This is a natural counterpart of the maximal invariant in the sequence

$$\widehat{W}_\varepsilon^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \frac{\Delta Y_t}{\hat{\sigma}_f}, \quad (2.6.1)$$

$$\widehat{W}_{\phi_g}^{(T)}(s) = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} s_t \sigma_g \phi_g \left(G^{-1} \left(\frac{T+1+R_t^+}{2(T+1)} \right) \right), \quad (2.6.2)$$

where (R_2^+, \dots, R_T^+) are the ranks of absolute values of $(\hat{\varepsilon}_2, \dots, \hat{\varepsilon}_T)$, and (s_2, \dots, s_T) are signs of $(\hat{\varepsilon}_2, \dots, \hat{\varepsilon}_T)$. The symmetric reference density g is assumed to be symmetric with variance σ_g , score function ϕ_g and quantile function G^{-1} . Under the symmetric density condition, $\widehat{W}_\varepsilon^{(T)}$ and $\widehat{W}_{\phi_f}^{(T)}$ weakly converges to W_ε and W_{ϕ_f} , respectively.

Remark 2.6.2 (Serial-correlated errors). In this remark, we discuss possible extension for the case where errors are possibly serially correlated. To be specific, in model equation (2.2.2), let v_t denote the innovation at time t instead of ε_t , and model it as $v_t = \gamma_1 v_{t-1} \cdots - \gamma_p v_{t-p} + \varepsilon_t$.¹¹ We assume the same assumptions on ε_t as above. The inference for ρ is adaptive to the presence of γ in the sense that their corresponding score functions are asymptotically independent (see Section 7 of Jansson (2008)). As one might expect, replacing γ by some consistent estimator $\hat{\gamma}$ will not affect the result of testing ρ (asymptotically). Recall the i.i.d error case considered above, we use ΔY_t , which actually plays the role of estimates $\hat{\varepsilon}_t$ for ε_t under the null hypothesis, and R_t , which is the rank of $\hat{\varepsilon}_t$, to build the HRT and the AHRT statistics. In this case, estimates for ε_t becomes $\hat{\varepsilon}_t = \Delta Y_t - \hat{\gamma}_1 \Delta Y_{t-1} \cdots - \hat{\gamma}_p \Delta Y_{t-p}$, and subsequently, the rank of the new estimates $\hat{\varepsilon}_t$, R_t , becomes aligned ranks. Consistency of $\hat{\gamma}$ gives the consistency of $\hat{\varepsilon}_t$. Then, with these consistent estimates of errors and associated aligned ranks, the convergences in (2.4.2) and (2.4.3) are preserved and hence the properties of HRTs and AHRTs are also preserved for the aligned-rank-based versions.

Remark 2.6.3 (Nonparametrically estimated reference density). The Hybrid Rank Test and the Approximate Hybrid Rank Test are optimal when the

¹¹General invertible ARMA process can be represented as an infinite order AR process.

reference density g coincides with the actual innovation density f . It is therefore reasonable to consider these test using a nonparametric estimate of f , say \hat{f} , as reference density. Commonly such estimate is based on the order statistics of the consistent innovation estimates, $\hat{\varepsilon}_t$, and thus independent of their ranks used in the HRT. Under a suitable consistency condition, the HRT based on \hat{f} asymptotically behaves as the HRT based on the true innovation density f . Thus, such test achieves the optimality properties of Theorem 2.4.1 and Theorem 2.4.2 globally. Notably, even if there exists relatively large bias in the estimation of f , the usage of rank statistics ensures zero expectation of the feasible score function $\phi_{\hat{f}}[\hat{F}^{-1}(R_t/(T+1))]$, which furthermore ensures the validity of the HRTs and the AHRTs. This argument can also be showed with the fact that rank-based score converges to Brownian bridge as $T \rightarrow \infty$, and $\int_0^1 W_\varepsilon(s)dB_{\phi_f}(s) = \int_0^1 \bar{W}_\varepsilon(s)dW_{\phi_f}(s)$, where $\bar{W}_\varepsilon(s) = W_\varepsilon(s) - \int_0^1 W_\varepsilon(s)ds$. Thus, a drift in W_{ϕ_f} caused by estimation bias will be canceled out.

2.6.1 Some simulation results

We also provide some simulation results for Remark 2.6.1 and Remark 2.6.2, with the corresponding modified inference strategies therein respectively mentioned. Since we mean to demonstrate the asymptotic properties, here we present the large sample performances with sample size $T = 2,500$.

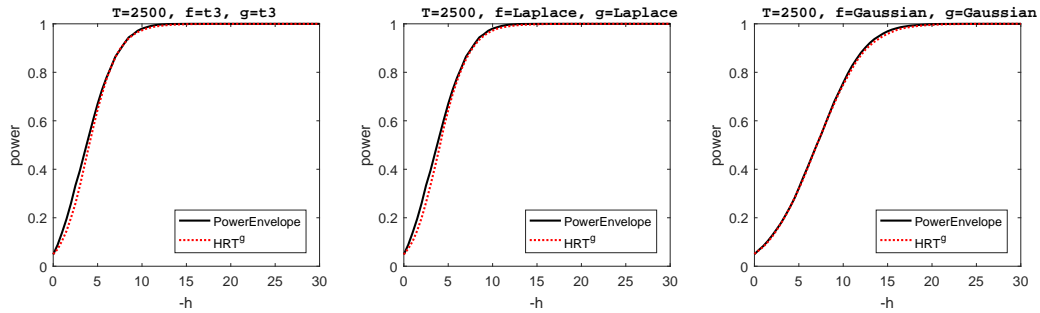


Figure 2.6: Parametric power envelopes and power functions of modified AHRT^g for symmetric density f case in Remark 2.6.1.

Under symmetric density assumption, Figure 2.6 shows that the power function of the modified Hybrid Rank Tests based on statistics in (2.6.1) and (2.6.2), with correctly specified reference density, can achieve the parametric power envelope. This numerically proves the adaptive result for the parameter ρ with respect to the unknown symmetric density function.

For serial-correlated error case, we employ the ARMA model on the error term v_t as below

$$\begin{aligned} Y_t &= \mu + X_t, \\ X_t &= \rho X_{t-1} + v_t, \\ v_t &= -0.5v_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}, \quad t \in \mathbb{N}, \end{aligned}$$

where the assumption on ε_t stay unchanged. For the inference procedure based on aligned ranks in Remark 2.6.2, we choose the AR regression order $p = 8$. For different combinations of true and reference densities, in Figure 2.7, we show that the power functions are of correct size and of similar power properties as the i.i.d error case. These results validate that the serial correlation in the errors can be well handled, as for the cases of other unit root tests, with an ancillary auto-regression on the increments of the observed process.

Small sample performances for these cases share similarities as those in Section 2.5.2, e.g., slightly lower size and power, comparing with their corresponding large sample performances. Simulation results for a variety of sample sizes are available upon request.

2.7 Conclusion

This paper has provided a structural representation of the limit experiment of the standard unit root model in a univariate but semiparametric setting. Using invariance arguments, we have derived the semiparametric power envelope. These invariance structures also lead, using the Neyman-Pearson lemma, to point-optimal semiparametric tests. The analysis naturally leads to the use of rank-based statistics.

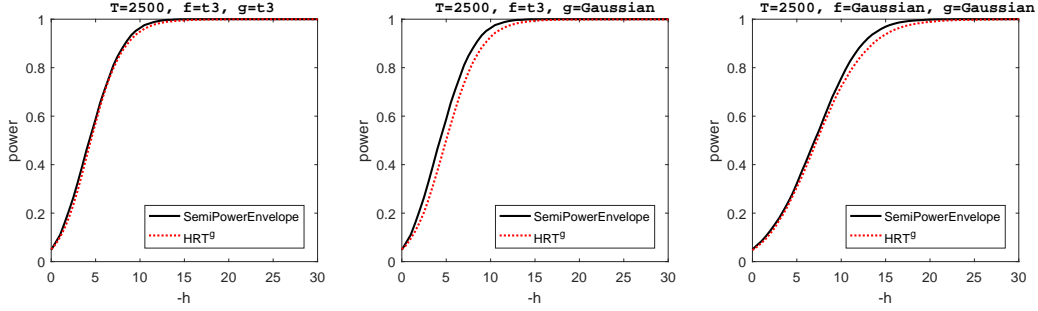


Figure 2.7: Semiparametric power envelopes and power functions of modified AHRT^g for serial-correlated error case in Remark 2.6.2.

Our tests are asymptotically valid, invariant, and (with a correctly chosen reference density) point-optimal. Moreover, we establish a Chernoff-Savage type property of our test: irrespective of the reference density chosen, our test outperforms its classical competitor which in this case is the ERS test. Finally, we introduced a simplified version of our test and show, in a small Monte-Carlo study, that our theoretical results carry over to small samples.

As potential future work we mention the use of similar ideas to construct hybrid rank-based tests in more general time-series models which, for instance, allow for serial correlation in the error terms, a deterministic time trend term, or stochastic volatility. Also, the structural representation of the limit experiment and its invariance properties can be applied to other non-stationary time-series models, for instance, cointegration and predictive regression models.

2.8 Appendix: Proofs

2.8.1 Preliminaries

Introduce the filtrations $\mathbb{F}^{(T)} := (\mathcal{F}_u^{(T)}, u \in [0, 1])$, $T \in \mathbb{N}$, defined by $\mathcal{F}_u^{(T)} := \sigma(Y_t, t \in \mathbb{N} : t \leq [uT])$, $u \in [0, 1]$. The angle-bracket process $\langle A_i^{(T)}, A_j^{(T)} \rangle(u)$ and the straight-bracket process $[A_i^{(T)}, A_j^{(T)}](u)$ are now

well-defined for all $\mathbb{F}^{(T)}$ -adapted locally square-integrable martingales and semimartingales $A_i^{(T)}$, respectively (see, e.g., Jacod and Shiryaev (2002)). If $A_i^{(T)}$, $i = 1, 2$, are square-integrable martingales of the form $A_i^{(T)}(u) = \sum_{t=1}^{[uT]} I_{Tt}^{(i)}$ with $I_{Tt}^{(i)}$ \mathcal{F}_t -measurable, we have

$$\begin{aligned} [A_1^{(T)}, A_2^{(T)}](u) &= \sum_{t=1}^{[uT]} I_{Tt}^{(1)} I_{Tt}^{(2)'} \\ \langle A_1^{(T)}, A_2^{(T)} \rangle(u) &= \sum_{t=1}^{[uT]} \mathbb{E}[I_{Tt}^{(1)} I_{Tt}^{(2)'} | \mathcal{F}_{t-1}]. \end{aligned}$$

Recall that for a square-integrable martingale with continuous sample paths the angle-brackets and straight-brackets coincide.

The lemma below shows that the partial sum processes introduced in Section 2.3.1 weakly converges to the associated Brownian motions. Due to the i.i.d.-ness of the innovations, the lemma is a direct corollary to the functional central limit theorem VIII.3.33 in Jacod and Shiryaev (2002).

Lemma 2.8.1. *Let $f \in \mathcal{F}$ and let, with $m \geq 3$, $k_1, \dots, k_{m-2} \in \mathbb{N}$. Define, with the notation of Section 2.3.1,*

$$\mathcal{W}^{(T)} = (W_\varepsilon^{(T)}, W_{\phi_f}^{(T)}, W_{b_1}^{(T)}, \dots, W_{b_{m-2}}^{(T)})'$$

and

$$\mathcal{W} = (W_\varepsilon, W_{\phi_f}, W_{b_1}, \dots, W_{b_{m-2}})'$$

Then, in $D_{\mathbb{R}^m}[0, 1]$ and under $P_{0,0;\mu,f}^{(T)}$, we have

$$\mathcal{W}^{(T)} \Rightarrow \mathcal{W}, \tag{2.8.1}$$

and, still under $P_{0,0;\mu,f}^{(T)}$,

$$\begin{aligned} \langle \mathcal{W}^{(T)}, \mathcal{W}^{(T)} \rangle(1) &= [\mathcal{W}^{(T)}, \mathcal{W}^{(T)}](1) + o_P(1) \\ &= \text{Var}_f(\mathcal{W}(1)) + o_P(1). \end{aligned} \tag{2.8.2}$$

2.8.2 Main proofs

Proof of Proposition 2.3.1.

For notational convenience we drop the superscript “ (T) ” in the following and thus write f_η instead of $f_\eta^{(T)}$. It is clear, since η has finite support, that we have $f_\eta > 0$ for large enough T . The mean restrictions $\int b_k(e)f(e)de = 0$, together with the finite support of η , guarantee that f_η integrates to 1. Similarly, $\int b_k(e)ef(e)de = 0$ implies $E_{f_\eta}[\varepsilon_t] = 0$. Of course, absolute continuity of f_η follows from $f \in \mathcal{F}$ and, again because η has finite support, $\sum_{k=1}^\infty \eta_k b \in C_{2,b}(\mathbb{R})$. These properties also easily yield $\text{Var}_{f_\eta}[\varepsilon_t] < \infty$. Only $J_{f_\eta} < \infty$ requires a bit of straightforward calculus. We have

$$f'_\eta(e) = f'(e) \left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^\infty \eta_k b_k(e) \right) + f(e) \frac{1}{\sqrt{T}} \sum_{k=1}^\infty \eta_k b'_k(e), \quad \text{a.e.}$$

There exist $C_1, C_2 < \infty$ such that we have (for all T) $\|1 + T^{-1/2} \sum_{k=1}^\infty \eta_k b_k\|_\infty \leq C_1$ and $\|T^{-1/2} \sum_{k=1}^\infty \eta_k b'_k\|_\infty^2 \leq C_2$. Moreover, there exists $C_3 > 0$ such that, for all $T \geq T'$, $\|(1 + T^{-1/2} \sum_{k=1}^\infty \eta_k b_k)^{-1}\|_\infty^2 \leq C_3$. Using these observations we immediately obtain, for $T \geq T'$,

$$\int \left(-\frac{f'_\eta(e)}{f_\eta(e)} \right)^2 f_\eta(e) de \leq 2C_1 J_f + 2C_2 C_3 \int f_\eta(e) de < \infty,$$

which concludes the proof. \square

Proof of Proposition 2.3.2.

We first note that Part (iii) directly follows from an application of Corollary 3.5.16 in Karatzas and Shreve (1991). In the following we evaluate, unless mentioned otherwise, expectations, O_P 's, and o_P 's under $P_{0,0;\mu,f}^{(T)}$. We first establish Part (ii) and prove the quadratic expansion (i) afterwards. Let k_1, \dots, k_m denote the elements for which η does not vanish, i.e. $\eta_{k_j} \neq 0$ and $\eta_i = 0$ for $i \notin \{k_1, \dots, k_m\}$. And let $a_T = (h_T, \eta_{k_1}, \dots, \eta_{k_m})'$ and $a = (h, \eta_{k_1}, \dots, \eta_{k_m})'$.

PROOF OF PART (ii): First we introduce auxiliary processes $\tilde{\Delta}^{(T)}$, $T \in \mathbb{N}$,

by

$$\tilde{\Delta}^{(T)}(r) = \left(\int_0^r W_\varepsilon^{(T)}(s-) dW_{\phi_f}^{(T)}(s), W_{b_{k_1}}^{(T)}(r), \dots, W_{b_{k_m}}^{(T)}(r) \right), r \in [0, 1].$$

A combination of Lemma 4.7.1 with Theorem 2.1 in Hansen (1992) (the conditions are trivially met) yields $\tilde{\Delta}^{(T)} \Rightarrow \tilde{\Delta}$ in $D_{\mathbb{R}^{m+1}}[0, 1]$, where $\tilde{\Delta}$ is given by

$$\tilde{\Delta}(r) = \left(\int_0^r W_\varepsilon(s) dW_{\phi_f}(s), W_{b_{k_1}}(r), \dots, W_{b_{k_m}}(r) \right)',$$

$r \in [0, 1]$ (which we evaluate under $\mathbb{P}_{0,0}$). Using this weak convergence, the identity $\langle A, B \rangle(r) = A(r)B(r) - A(0)B(0) - \int_0^r A(s-)dB(s) - \int_0^r B(s-)dA(s)$ ¹² and the continuous mapping theorem in combination with Theorem 2.1 in Hansen (1992) (the condition to this theorem is met as $\tilde{\Delta}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$ and as we have $\sum_{t=1}^T \mathbb{E}|\tilde{\Delta}^{(T)}(t/T) - \tilde{\Delta}^{(T)}((t-1)/T)|^2 = O(1)$) yields,

$$\left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right] (1) \right) \Rightarrow \left(\tilde{\Delta}(1), \langle \tilde{\Delta}, \tilde{\Delta} \rangle (1) \right). \quad (2.8.3)$$

The quadratic variation at time 1, $\langle \tilde{\Delta}, \tilde{\Delta} \rangle_1$, is given by

$$\begin{aligned} \langle \tilde{\Delta}_1, \tilde{\Delta}_1 \rangle_1 &= J_f \int_0^1 W_\varepsilon^2(s) ds, \quad \langle \tilde{\Delta}_{j+1}, \tilde{\Delta}_{j+1} \rangle_1 = 1, \\ \langle \tilde{\Delta}_1, \tilde{\Delta}_{j+1} \rangle_1 &= J_{k_j, f} \int_0^1 W_\varepsilon(s) ds, \text{ and } \langle \tilde{\Delta}_{1+i}, \tilde{\Delta}_{1+j} \rangle_1 = 0, \end{aligned}$$

or $i \neq j \in \{1, \dots, m\}$. The angle brackets of $\tilde{\Delta}^{(T)}$ at time 1, $\langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \rangle_1$, take a similar form (just replace the limiting Brownian motions by their empirical analogues). On noting that we have

$$\mathcal{I}_f^{(T)}(h, \eta) = a' \langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \rangle_1 a \quad \text{and} \quad \mathcal{I}_f(h, \eta) = a' \langle \tilde{\Delta}, \tilde{\Delta} \rangle_1 a, \quad (2.8.4)$$

the proof of Part (ii) follows if we show

$$\langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \rangle_1 - \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right]_1 = o_P(1). \quad (2.8.5)$$

¹²With continuous argument, the angle-bracket $\langle \rangle_t$ (or $\langle \rangle(t)$) stands for the quadratic variation at time t .

Indeed, a combination of (2.8.3)-(2.8.5) with the continuous mapping theorem yields (ii). To demonstrate (2.8.5), we first note that Lemma 4.7.1 yields, for $i, j = 1, \dots, m$,

$$\left\langle \tilde{\Delta}_{1+i}^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right\rangle (1) - \left[\tilde{\Delta}_{1+i}^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right] (1) = o_P(1).$$

So we only need to consider

$$\begin{aligned} r_1^{(T)} &= \left[\tilde{\Delta}_1^{(T)}, \tilde{\Delta}_1^{(T)} \right]_1 - \left\langle \tilde{\Delta}_1^{(T)}, \tilde{\Delta}_1^{(T)} \right\rangle_1 \\ &= \frac{1}{T^2} \sum_{t=2}^T \sigma_f^{-2} (Y_{t-1} - Y_1)^2 (\sigma_f^2 \phi_f^2(\varepsilon_t) - J_f) \quad \text{and} \\ r_{2,j}^{(T)} &= \left[\tilde{\Delta}_1^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right]_1 - \left\langle \tilde{\Delta}_1^{(T)}, \tilde{\Delta}_{1+j}^{(T)} \right\rangle_1 \\ &= \frac{1}{T^{3/2}} \sum_{t=2}^T \sigma_f^{-1} (Y_{t-1} - Y_1) (\sigma_f \phi_f(\varepsilon_t) b_{k_j}(\varepsilon_t) - J_{f,k_j}), \end{aligned}$$

for $j = 1, \dots, m$. We have

$$\mathbb{E}(r_{2,j}^{(T)})^2 = \frac{1}{T} \text{Var}_f [\sigma_f \phi_f(\varepsilon_1) b_{k_1}(\varepsilon_t)] \int_0^1 \mathbb{E}(W_\varepsilon^{(T)}(u-))^2 du = o(1).$$

For $r_1^{(T)}$ the same line of reasoning can be followed in case $\phi_f(\varepsilon_1)$ has a finite fourth moment. This, however, does not need to be the case under Assumption 2.4.1. Therefore we resort to an application of Theorem 2.23 in Hall and Heyde (1980) which shows that $r_1^{(T)} = o_P(1)$ if, for all $\delta > 0$,

$$\sum_{t=2}^T \frac{1}{T^2} \mathbb{E} \left[(Y_{t-1} - Y_1)^2 \phi_f^2(\varepsilon_t) 1_{\{|(Y_{t-1} - Y_1) \phi_f(\varepsilon_t)| > \delta T\}} \mid \mathcal{F}_{t-1} \right] = o_P(1). \quad (2.8.6)$$

Using the notation $\zeta(M) = \mathbb{E} \left[\sigma_f^2 \phi_f^2(\varepsilon_1) 1_{\{|\sigma_f \phi_f(\varepsilon_1)| \geq M\}} \right]$, we see that the left-hand-side of the previous display is bounded by

$$\zeta \left(\frac{\delta \sqrt{T}}{\|W_\varepsilon^{(T)}\|_\infty} \right) \int_0^1 \left(W_\varepsilon^{(T)}(u-) \right)^2 du = o_P(1),$$

by a combination of Lemma 4.7.1, the continuous mapping theorem, and $\zeta(M) \rightarrow 0$ as $M \rightarrow \infty$ (dominated convergence). This concludes the proof of Part (ii).

PROOF OF PART (i): We use Proposition 1 in Hallin et al. (2015) to prove the expansion. To this end we set $\tilde{P}_T = P_{h_T, \eta; \mu, f}^{(T)}$, $P_T = P_{0,0; \mu, f}^{(T)}$, and $\mathcal{F}_{Tt} = \sigma(Y_1, \dots, Y_t)$. And we introduce

$$S_{Tt} = \left(T^{-1}(Y_{t-1} - Y_1) \phi_f(\Delta Y_t), T^{-1/2} b_{k_1}(\Delta Y_t), \dots, T^{-1/2} b_{k_m}(\Delta Y_t) \right)'$$

for $t = 2, \dots, T$ and $T \in \mathbb{N}$. Notice that (see the proof of Part (ii) above) $\tilde{\Delta}^{(T)}(1) = \sum_{t=1}^T S_{Tt}$.

In the notation of Proposition 1 in Hallin et al. (2015) we have, for $t \geq 2$,

$$LR_{Tt} = \frac{f(\Delta Y_t - w_{Tt}) \left(1 + \frac{1}{\sqrt{T}} \sum_{j=1}^m \eta_{k_j} b_{k_j}(\Delta Y_t - w_{Tt}) \right)}{f(\Delta Y_t)} \quad (2.8.7)$$

with

$$w_{Tt} = \frac{h_T}{T} (Y_{t-1} - \mu).$$

Assumption 2.2.1 implies (see, e.g., Le Cam (1986, Section 17.3) and Le Cam and Yang (2000, Section 7.3)) that the mapping $e \mapsto f^{1/2}(e)$ is differentiable in quadratic mean:

$$\frac{\sqrt{f}(e - w)}{\sqrt{f}(e)} = 1 + \frac{1}{2} [\phi_f(e)w + r(e, w)],$$

where

$$Er^2(\varepsilon_1, w) = o_P(w^2), \quad (2.8.8)$$

which implies, by Cauchy-Schwarz inequality,

$$Er(\varepsilon_1, w) = o_P(w). \quad (2.8.9)$$

Let $B_{Tt} = T^{-1/2} \sum_{j=1}^m \eta_{k_j} b_{k_j}(\Delta Y_t - w_{Tt})$, $B_{Tt}^0 = T^{-1/2} \sum_{j=1}^m \eta_{k_j} b_{k_j}(\Delta Y_t)$, and introduce

$$R_{Tt}^b = 2 \left(\sqrt{1 + B_{Tt}} - 1 - \frac{1}{2} B_{Tt}^0 \right), \quad (2.8.10)$$

where, by Taylor's theorem (twice) and the assumption that the continuous derivatives of b_{k_j} are bounded, we have

$$\max_{2 \leq t \leq T} |R_{Tt}^b| = o_P\left(\frac{1}{T}\right). \quad (2.8.11)$$

Recall that $a_T = (h_T, \eta_{k_1}, \dots, \eta_{k_m})'$ and $a = (h, \eta_{k_1}, \dots, \eta_{k_m})'$. We have, for $t \geq 2$,

$$\begin{aligned} \sqrt{LR_{Tt}} &= \left(1 + \frac{1}{2}w_{Tt}\phi_f(\varepsilon_t) + \frac{1}{2}r(\varepsilon_t, w_{Tt})\right) \left(1 + \frac{1}{2}B_{Tt}^0 + \frac{1}{2}R_{Tt}^b\right) \\ &= 1 + \frac{1}{2}a'_T S_{Tt} + \frac{1}{2}R_{Tt}, \end{aligned}$$

with

$$R_{Tt} = r(\varepsilon_t, w_{Tt}) + R_{Tt}^b + \frac{1}{2}(B_{Tt}^0 + R_{Tt}^b)(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})). \quad (2.8.12)$$

So we can conclude that expansion (i) holds once we verify the conditions in Proposition 1 of Hallin et al. (2015).

Condition (a). This is immediate as a_T converges by assumption.

Condition (b). Square-integrability follows from our assumption $f \in \mathcal{F}$. Display (2) in Condition (b) of Hallin et al. (2015) follows immediately from the independence of ε_t and $\mathcal{F}_{T,t-1}$, $E\phi_f(\varepsilon_t) = 0$, and $Eb_{k_j}(\varepsilon_t) = 0$, $j = 1, \dots, m$. The second equation in Display (3) in Condition (b) is immediate as $J_T = \sum_{t=1}^T E[S_{Tt}S'_{Tt}|\mathcal{F}_{t-1}] = \langle \tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \rangle_1 = O_P(1)$ (see (2.8.3)). Next we verify the conditional Lindeberg condition (the first equation in Display (3)), which is, for all $\delta > 0$,

$$\sum_{t=2}^T E\left[(a'_T S_{Tt})^2 1_{\{|a'_T S_{Tt}| > \delta\}}|\mathcal{F}_{t-1}\right] = o_P(1). \quad (2.8.13)$$

Observe

$$\begin{aligned} &\sum_{t=2}^T E\left[(a'_T S_{Tt})^2 1_{\{|a'_T S_{Tt}| > \delta\}}|\mathcal{F}_{t-1}\right] \\ &= \sum_{t=2}^T E\left[\left(\frac{h_T}{T}(Y_{t-1} - Y_1)\phi_f(\Delta Y_t) + \sum_{j=1}^m \frac{\eta_{k_j}}{\sqrt{T}}b_{k_j}(\Delta Y_t)\right)^2 1_{\{(a'_T S_{Tt})^2 > \delta^2\}}|\mathcal{F}_{t-1}\right] \end{aligned}$$

$$\begin{aligned} \leq & (m+1)^2 \sum_{t=2}^T \mathbb{E} \left[\left(\frac{h_T^2}{T^2} (Y_{t-1} - Y_1)^2 \phi_f^2(\Delta Y_t) 1_{\{(m+1)^2 h_T^2 (Y_{t-1} - Y_1)^2 \phi_f^2(\Delta Y_t) > \delta^2 T^2\}} \right) \middle| \mathcal{F}_{t-1} \right] \\ & + \sum_{j=1}^m (m+1)^2 \sum_{t=2}^T \mathbb{E} \left[\frac{\eta_{k_j}^2}{T} b_{k_j}^2(\Delta Y_t) 1_{\{(m+1)^2 \eta_{k_j}^2 b_{k_j}^2(\Delta Y_t) > \delta^2 T\}} \middle| \mathcal{F}_{t-1} \right]. \end{aligned}$$

To complete the proof, we just need to show separately that, for any given $\delta > 0$, it holds that

$$\sum_{t=2}^T \mathbb{E} \left[\left(\frac{h_T^2}{T^2} (Y_{t-1} - Y_1)^2 \phi_f^2(\Delta Y_t) 1_{\{|(m+1)h_T(Y_{t-1} - Y_1)\phi_f(\Delta Y_t)| > \delta T\}} \right) \middle| \mathcal{F}_{t-1} \right] = o_P(1),$$

and

$$\sum_{t=2}^T \mathbb{E} \left[\frac{\eta_{k_j}^2}{T} b_{k_j}^2(\Delta Y_t) 1_{\{|(m+1)\eta_{k_j} b_{k_j}(\Delta Y_t)| > \delta \sqrt{T}\}} \middle| \mathcal{F}_{t-1} \right] = o_P(1).$$

Here, these two equalities can be shown in the same way as (2.8.6) is proved.

Condition (c). By (4.7.6) and (2.8.9), we have $\mathbb{E}[r^2(\varepsilon_t, w_{Tt}) | \mathcal{F}_{t-1}] = o_P(T^{-2})$ and $\mathbb{E}[r(\varepsilon_t, w_{Tt}) | \mathcal{F}_{t-1}] = o_P(T^{-1})$. Moreover, since all b_{k_j} s are bounded, we have $\max_{2 \leq t \leq T} |B_{Tt}^0| = O_P(T^{-1/2})$ and $\sum_{t=1}^T \mathbb{E}[B_{Tt}^0 | \mathcal{F}_{t-1}] = O_P(1)$. Together with (2.8.11), this yields

$$\begin{aligned} \sum_{t=2}^T \mathbb{E} \left[(r(\varepsilon_t, w_{Tt}) + R_{Tt}^b)^2 \middle| \mathcal{F}_{t-1} \right] &= o_P(1), \\ \sum_{t=2}^T \mathbb{E} \left[(r(\varepsilon_t, w_{Tt}) + R_{Tt}^b)(B_{Tt}^0 + R_{Tt}^b)(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})) \middle| \mathcal{F}_{t-1} \right] &= o_P(1), \\ \sum_{t=2}^T \mathbb{E} \left[\left((B_{Tt}^0 + R_{Tt}^b)(\phi_f(\varepsilon_t)w_{Tt} + r(\varepsilon_t, w_{Tt})) \right)^2 \middle| \mathcal{F}_{t-1} \right] &= o_P(1), \end{aligned}$$

and, thus, by (2.8.12), we have

$$\sum_{t=2}^T \mathbb{E} [R_{Tt}^2 | \mathcal{F}_{t-1}] = o_P(1).$$

This establishes Display (4). As we assumed the density f to be strictly positive, Display (5) is immediate by plugging in (2.8.7) to its left-hand side.

Condition (d). This follows easily from

$$\log \frac{f_\eta^{(T)}(Y_1 - \mu)}{f(Y_1 - \mu)} = \log \left[1 + \frac{1}{\sqrt{T}} \sum_{j=1}^m \eta_{k_j} b_{k_j}(\varepsilon_1) \right] = o_P(1).$$

This completes the proof. \square

Proof of Theorem 2.3.1.

Let G be the group of translations g_η with $\eta \in c_{00}$ defined in (2.3.7). Invariance of M has been shown in Section 2.3.3. In order to prove that M is maximal invariant, following the idea in Section 6.2 of Lehmann and Romano (2005), we establish that $M((W_\varepsilon(s), (B_{b_k}(s))_{k \in \mathbb{N}})', s \in [0, 1]) = M((\widetilde{W}_\varepsilon(s), (\widetilde{B}_{b_k}(s))_{k \in \mathbb{N}})', s \in [0, 1])$ implies $\widetilde{W}_\varepsilon(s) = W_\varepsilon(s)$ and $\widetilde{W}_{b_k}(s) = g_{\eta_k}(W_{b_k}(s))$ with $s \in [0, 1]$, for some $g_\eta = (g_{\eta_k})_{k \in \mathbb{N}} \in G$.

Suppose $M((W_\varepsilon(s), (B_{b_k}(s))_{k \in \mathbb{N}})', s \in [0, 1]) = M((\widetilde{W}_\varepsilon(s), (\widetilde{B}_{b_k}(s))_{k \in \mathbb{N}})', s \in [0, 1])$, that is

$$\begin{aligned} W_\varepsilon(s) &= \widetilde{W}_\varepsilon(s), \\ B_{b_k}(s) &= \widetilde{B}_{b_k}(s), \quad k \in \mathbb{N}. \end{aligned}$$

This implies, for $\eta_k = W_{b_k}(1) - \widetilde{W}_{b_k}(1)$,

$$\begin{aligned} W_\varepsilon(s) - \widetilde{W}_\varepsilon(s) &= 0, \\ W_{b_k}(s) - \widetilde{W}_{b_k}(s) &= \eta_k s, \quad k \in \mathbb{N}. \end{aligned}$$

Hence $\widetilde{W}_\varepsilon(s) = W_\varepsilon(s)$ and $\widetilde{W}_{b_k}(s) = g_{\eta_k}(W_{b_k}(s))$ with $s \in [0, 1]$, which completes the proof. \square

Proof of Proposition 2.4.1.

Recall $\mathcal{M}_g = \sigma(W_\varepsilon, B_{\phi_g}) \subseteq \mathcal{M} = \sigma(W_\varepsilon, B_{b_k}; k \in \mathbb{N})$, the likelihood ratio of \mathcal{M}_g can be derived by taking the expectation of the likelihood ratio of \mathcal{M} conditional on the information \mathcal{M}_g . We find

$$\begin{aligned} \frac{d\mathbb{P}_h^{\mathcal{M}_g}}{d\mathbb{P}_0^{\mathcal{M}_g}} &= \mathbb{E}_0 \left[\frac{d\mathbb{P}_h^{\mathcal{M}}}{d\mathbb{P}_0^{\mathcal{M}}} \middle| \mathcal{M}_g \right] \\ &= \mathbb{E}_0 \left[\exp \left(h \left\{ \int_0^1 W_\varepsilon(s) dB_{\phi_f}(s) + W_\varepsilon(1) \int_0^1 W_\varepsilon(s) ds \right\} \right) \right] \end{aligned}$$

$$-\frac{1}{2}h^2 \left\{ J_f \int_0^1 W_\varepsilon^2(s)ds - \left(\int_0^1 W_\varepsilon(s)ds \right)^2 (J_f - 1) \right\} \Big| \mathcal{M}_g \Big].$$

Based on the covariance matrix (4.2.15) and using $\lambda = (J_{fg}\sigma_{\varepsilon\phi_g} - \sigma_{\varepsilon\phi_g}^2)/(J_g - \sigma_{\varepsilon\phi_g}^2)$, we have the decomposition $W_{\phi_f} = (1 - \lambda)W_\varepsilon + \lambda W_{\phi_g}/\sigma_{\varepsilon\phi_g} + W_\dagger$, where W_\dagger is a Brownian motion (not necessarily standard) independent of both W_ε and W_{ϕ_g} . Together with the decomposition $W_{\phi_g}/\sigma_{\varepsilon\phi_g} = W_\varepsilon + \sqrt{J_g/\sigma_{\varepsilon\phi_g}^2 - 1}W_\perp$, we have

$$W_{\phi_f} = W_\varepsilon + \lambda \sqrt{\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1} W_\perp + W_\dagger.$$

Define $B_\varepsilon = B^{W_\varepsilon}$, $B_\perp = B^{W_\perp}$, and $B_\dagger = B^{W_\dagger}$. It follows

$$B_{\phi_f} = B_\varepsilon + \lambda \sqrt{\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1} B_\perp + B_\dagger.$$

Plugging this into the previous equation leads to

$$\begin{aligned} & \frac{d\mathbb{P}_h^{\mathcal{M}_g}}{d\mathbb{P}_0^{\mathcal{M}_g}} \\ &= \mathbb{E}_0 \left[\exp \left(h \left\{ \int_0^1 W_\varepsilon(s) \left(dB_\varepsilon(s) + \lambda \sqrt{\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1} dB_\perp(s) + dB_\dagger(s) \right) \right. \right. \right. \\ & \quad \left. \left. \left. + W_\varepsilon(1) \int_0^1 W_\varepsilon(s)ds \right\} - \frac{1}{2}h^2 \left\{ J_f \int_0^1 W_\varepsilon^2(s)ds - \left(\int_0^1 W_\varepsilon(s)ds \right)^2 (J_f - 1) \right\} \right\} \Big| \mathcal{M}_g \right] \\ &= \mathbb{E}_0 \left[\exp \left(h \left\{ \int_0^1 W_\varepsilon(s) dW_\varepsilon(s) + \lambda \sqrt{\frac{J_g}{\sigma_{\varepsilon\phi_g}^2} - 1} \int_0^1 W_\varepsilon(s) dB_\perp(s) + \int_0^1 W_\varepsilon(s) dB_\dagger(s) \right\} \right. \right. \\ & \quad \left. \left. - \frac{1}{2}h^2 \left\{ J_f \int_0^1 W_\varepsilon^2(s)ds - \left(\int_0^1 W_\varepsilon(s)ds \right)^2 (J_f - 1) \right\} \right\} \Big| \mathcal{M}_g \right] \\ &= \mathbb{E}_0 \left[\exp \left(h \{ \Delta_\varepsilon + \lambda \Delta_\perp + \Delta_\dagger \} - \frac{1}{2}h^2 \{ \langle \Delta_\varepsilon + \lambda \Delta_\perp + \Delta_\dagger, \Delta_\varepsilon + \lambda \Delta_\perp + \Delta_\dagger \rangle \} \right) \Big| \mathcal{M}_g \right] \end{aligned}$$

where Δ_ε and Δ_\perp are defined in the present proposition, and $\Delta_\dagger := \int_0^1 W_\varepsilon(s) dB_\dagger(s)$.

Under $\mathbb{P}_{0,0}$ the process B_\dagger is independent of W_ε and B_{ϕ_g} (henceforth B_\perp).

Consequently, $\langle \Delta_\varepsilon, \Delta_\dagger \rangle = 0$ and $\langle \Delta_\perp, \Delta_\dagger \rangle = 0$. Noting that Δ_ε , Δ_\perp , $\langle \Delta_\varepsilon \rangle$, $\langle \Delta_\perp \rangle$ and $\langle \Delta_\dagger \rangle$ are all \mathcal{M}_g -measurable, we thus obtain,

$$\frac{d\mathbb{P}_h^{\mathcal{M}_g}}{d\mathbb{P}_0^{\mathcal{M}_g}} = \mathbb{E}_0 \left[\exp \left(h \{ \Delta_\varepsilon + \lambda \Delta_\perp + \Delta_\dagger \} - \frac{1}{2}h^2 \{ \langle \Delta_\varepsilon \rangle + \lambda^2 \langle \Delta_\perp \rangle + \langle \Delta_\dagger \rangle \} \right) \Big| \mathcal{M}_g \right]$$

$$\begin{aligned}
&= \exp \left(h \{ \Delta_\varepsilon + \lambda \Delta_\perp \} - \frac{1}{2} h^2 \{ \langle \Delta_\varepsilon \rangle + \lambda^2 \langle \Delta_\perp \rangle + \langle \Delta_\dagger \rangle \} \right) \mathbb{E}_0 \left[\exp (h \Delta_\dagger) \mid \mathcal{M}_g \right] \\
&= \exp \left(h \{ \Delta_\varepsilon + \lambda \Delta_\perp \} - \frac{1}{2} h^2 \{ \langle \Delta_\varepsilon \rangle + \lambda^2 \langle \Delta_\perp \rangle \} \right).
\end{aligned}$$

The last equality holds since $\mathbb{E}_0 \left[\exp (h \Delta_\dagger) \mid \mathcal{M}_g \right] = \exp \left(\frac{1}{2} h^2 \langle \Delta_\dagger \rangle \right)$. \square

Chapter 3

A New Semiparametric Approach for Nonstandard Econometric Problems

Abstract. This paper develops a new approach to derive efficiency bounds and efficient tests for nonstandard problems in semiparametric econometric models where the innovation density is treated as an infinite-dimensional nuisance parameter. Our approach is based on an explicit nonparametric modeling of the innovation density and a structural version of the limit experiment. The structural limit experiment exhibits an invariance restriction that we use to eliminate the nuisance parameter. The associated maximal invariant gives the efficiency bound using Neyman-Pearson lemma. Moreover, the invariance structure naturally leads to statistical inference procedures. For example, the appearance of Brownian Bridges suggests the use of rank statistics. A simple linear regression model is used as a running example, leading to the same efficiency bounds and tests as the traditional least-favorable parametric submodel approach. We apply our results to two nonstandard econometric problems:

testing cointegration rank and testing problem with weak instruments. In both cases, we derive the semiparametric power envelopes and the induced rank-based inference procedures.

Key words. Semiparametric efficiency, nonstandard econometric problems, structural limit experiment, maximal invariant, rank statistic, cointegration, weak instrument.

3.1 Introduction

Semiparametric efficiency issue has been studied for decades in the econometrics and statistics literature. The theory for standard problems in semiparametric econometric models is well developed for both estimation and testing procedures (e.g., Newey (1990), Bickel et al. (1998), Van der Vaart (2000)). However, it is far from trivial to extend the existing tools to nonstandard econometric problems.¹ The development of a general framework for such cases is still ongoing. For instance, Jansson (2008) derived the semiparametric power envelopes using the canonical *least-favorable parametric submodel* (LFPS) approach for the unit root testing problem. Using an alternative approach, Zhou, van den Akker, and Werker (2016) rederived this semiparametric power envelope for asymptotic invariant tests but, inspired by the invariant structure, they also provided a new inference procedure based on rank statistics — the Hybrid Rank Tests — for unit root testing.

As a natural subsequent step, in Section 3.2, I provide a general framework of this approach for a semiparametric econometric models in which the innovation density is taken as an infinite-dimensional nuisance parameter. Here the likelihood ratio admits the *locally asymptotically Brownian functional* (LABF) form in the sense of Jeganathan (1995). We take *locally asymptotically normal* (LAN) and *locally asymptotically mixed normal* (LAMN) models as special cases of the LABF form.² Unlike the traditional

¹The concept of nonstandard econometric problem refers to Müller and Norets (2016).

²The more general concept of *Locally Asymptotically Quadratic* (LAQ) is not employed

LFPS approach that begins with some parametric submodels embedding the true unknown model (and finds the least-favorable one), in our approach, we start with an explicit nonparametric modeling of the density f . In this (non-parametric) model, we employ an orthonormal basis in a suitable functional space as perturbation functions, and an infinite-dimensional local parameter η to describe the deviation from the true density. Then, using standard techniques, we obtain the LABF likelihood ratio under mild conditions. As the likelihood ratio can also be regarded as a Radon-Nikodym derivative, an application of the Girsanov's Theorem leads to a structural version of this limit experiment. We call this the "structural limit experiment".

This structural limit experiment is key to our approach of exploiting invariance structures. In the structural limit experiment, driven by LABF type likelihood ratios, we observe an infinite-dimensional process. It is a Brownian motion under the null hypothesis, while, under the alternative hypothesis, it takes the form of an Ornstein-Uhlenbeck (OU) process. The nuisance parameter η appears in the drift term. This feature suggests to impose the invariance restriction to eliminate η . To be specific, taking the bridges of a process removes the drift term and, as a consequence, the bridge process obtained is invariant with respect to (w.r.t.) any drift transformation in the original process. Applying this operation to all the elements (of the infinite-dimensional observation process) that are affected by η , we can get an invariant sigma-field. Even more, we prove that it is *maximally invariant*. This result is meaningful in the sense that any invariant statistic must be a function of this maximal invariant. Consequently, by the Neyman-Pearson lemma and the Asymptotic Representation Theorem, the power of the test based on the maximal invariant's likelihood ratio provides an upper bound for the power of all asymptotically invariant tests.

For standard LAN cases, results derived by the new approach should coincide with existing ones in the literature. We consider the linear regression

here in case some unanticipated problems have log-likelihood ratios of LAQ form but not of the LABF form.

model as a running example. In the traditional LAN limit experiment, we observe one draw from a normally distributed random variable. However, in the structural limit experiment, this random variable is actually a stochastic integral of a deterministic integrand with respect to an OU process. After eliminating the nuisance parameter η in the way described above, we find that maximal invariant's likelihood ratio induces the *efficient score function*. This concept, as well as our new approach, can be smoothly extended to all LABF models (including LAN and LAMN).

Furthermore, we also use the structural limit experiment and its invariance structures to suggest concrete inference procedures. In the LABF model, the efficient score contains a stochastic integral of the form $\int_0^1 M(u, h) dB_{\ell_f}(u)$, where B_{ℓ_f} is the Brownian bridge of a given Brownian motion W_{ℓ_f} . A simple derivation shows that this term also equals $\int_0^1 \widetilde{M}(u, h) dW_{\ell_f}(u)$ with $\widetilde{M}(u, h) = M(u, h) - \int_0^1 M(u, h) du$. Both expressions are reminiscent of two strands in the statistical literature: inference based on a nonparametrically estimated f and rank-based inference.

- (1) **\hat{f} -based inference associated with $\int_0^1 \widetilde{M}(u, h) dW_{\ell_f}(u)$:** the appearance of $\widetilde{M}(u, h)$ enables the use of an estimated density \hat{f} even though f cannot be estimated \sqrt{T} -unbiasedly (T denotes the sample size) when it is unrestricted. However, any bias in the estimation will be canceled out by the centered integrand \widetilde{M} and, thus, validity of the inference is preserved.³
- (2) **Rank-based inference associated with $\int_0^1 M(u, h) dB_{\ell_f}(u)$:** the Brownian bridge B_{ℓ_f} naturally suggests inference based on rank statistics. The reason for this is that the partial sum of scores using the innovations' ranks and based on a reference density g converges to a Brownian bridge. This Brownian bridge equals B_{ℓ_f} when $g = f$, establishing the efficiency at correctly specified reference density.

³More explicitly, this is due to the fact that, for any value of bias in the limit denoted by a , we have $\int_0^1 \widetilde{M}(u, h) d(au + W_{\ell_f}(u)) = \int_0^1 \widetilde{M}(u, h) dW_{\ell_f}(u) + a \int_0^1 \widetilde{M}(u, h) du = \int_0^1 \widetilde{M}(u, h) dW_{\ell_f}(u)$.

These two techniques have different features. Inference based on a consistent estimate \hat{f} is semiparametrically and asymptotically optimal. On the other hand, the rank-based inference is locally optimal at g .⁴ However, the latter one enjoys some other desirable properties, e.g., an exact size in LAN models, Chernoff-Savage (type) result in LAN and LABF models, and it is easier to implement.

One of the challenges in the present paper is to extend these results from the univariate case to the multivariate case for rank-based inference. This is because vectors in \mathbb{R}^d , $d \geq 2$, are not naturally ordered. One common approach in the literature is to assume that f is elliptical, which allows these vectors ordered to be by the Mahalanobis distance. However, this may still be somewhat restrictive in many applications. To deal with this situation, in Section 3.3, we propose an inference procedure based on component rank statistics, the score structure of the multivariate normal distribution and d arbitrary marginal reference densities. In this way, no more assumptions (than the ones used to establish the LABF result) are needed. Simulation results show that this approach can achieve substantial efficiency gains when the innovations are “far from” normal distributed, e.g., multivariate Student’s t_3 distributed.

We apply our new approach to two nonstandard problems: testing the cointegration rank and testing problem with weak instruments. In the cointegration application, we firstly develop the semiparametric power envelope for asymptotically invariant tests, which is shown to be sharp with an example \hat{f} -based test. Then, based on the rank statistic proposed in Section 3.3, we also propose a family of rank-based tests.

For the weak instruments problem, semiparametric efficiency is studied in Cattaneo, Crump, and Jansson (2012). The authors show that the inference procedure is adaptive in the sense that it can proceed “as if” f is known, by a direct application of Bickel (1982)’s result for a regression

⁴This can be fixed by choosing $g = \hat{f}$, though this makes the inference procedure harder to implement.

model with zero mean regressors. Also, they give the \hat{f} -based versions of the AR test by Anderson, and Rubin (1949), the LM test by Kleibergen (2002), and the CLR test by Moreira (2003). Their analysis is under the assumption that the slope parameter under the null hypothesis, β_0 , is zero. Even though this assumption causes no efficiency loss, we relax it in order to reveal the asymptotic implications of the AR, LM, and CLR tests in our structural limit experiment. With β_0 being unrestricted, we reclaim the adaptivity result by showing that the semiparametric efficient score coincides the parametric one. As for the inference procedure, \hat{f} -based versions of the AR, LM, and CLR tests are provided in Cattaneo, Crump, and Jansson (2012). On the other strand, Andrews, and Marmer (2008) and Andrews, and Soares (2007) introduce rank-based versions, which, however, are only based on the ranks of the innovations of the first equation. This makes the rank-based LM and CLR tests (i) lose the exact size property of rank tests in LAN models and, (ii) gives up the information coming from the density of the other innovation series. To cover these aspects, we provided new rank-based versions of the AR, LM, and CLR tests, which are of exact (finite-sample) size and can gain efficiency using the joint density of the innovations.

The remainder of the paper is organized as follows. Section 3.2 introduces the general framework of the new approach. To illustrate the approach, a simple regression model is taken as a running example. Section 3.3 proposes an procedure for constructing test statistics based on multivariate componentwise ranks. The applications to cointegration and weak instruments are given in Section 3.4 and 3.5, respectively. Simulation studies for these two applications are provided in the above two sections separately. Section 3.6 concludes. Proofs are collected in Appendix 3.7, while calculations of conditional expectation on maximal invariants for all cases are collected in Appendix 3.8.

3.2 The New Approach

This section formalizes our new approach. We set up the model in Section 3.2.1 together with some necessary preliminaries for the asymptotic analysis. Section 3.2.2 introduces the LABF form of the log-likelihood ratios, based on which the structural limit experiment is derived in Section 3.2.3. Using this structural limit experiment, Section 3.2.4 gives the maximal invariant and the induced efficiency bounds. Section 3.2.5 discusses the resulting semiparametrically efficient inference procedures. Some additional discussions are collected in Section 3.2.6.

3.2.1 Model Setup

Suppose, in a sample of size $T \in \mathbb{N}$, we observe $Y_T = (y_1, \dots, y_T)' \in \mathbb{R}^{T \times p}$ for the dependent variable, and $X_T = (x_1, \dots, x_T)' \in \mathbb{R}^{T \times q}$ and $Z_T = (z_1, \dots, z_T)' \in \mathbb{R}^{T \times k}$ as two potential groups of nonrandom explanatory variables. Consider a model specified by a known nonrandom function ϕ ,

$$y_t = \phi(y_{t-1}, x_t, z_t; \theta) + \mu + \varepsilon_t, \quad (3.2.1)$$

where $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{p,t})' \in \mathbb{R}^p$, $t = 1, \dots, T$, are zero-mean i.i.d. innovations with density f , and the parameter of interest θ spanning in the parameter space $\Theta \subseteq \mathbb{R}^d$. The parameter μ stands for the mean of error term (or, the location of the density f). We impose the following mild assumption on f .

Assumption 3.2.1 (Innovation density f). (a) The p -dimensional density

f is absolutely continuous with a.e. derivative $\dot{f} = (\dot{f}_1, \dots, \dot{f}_p)$.

(b) The Fisher-information for location, $J_f = (J_{f,ij}) = \mathbb{E}[\ell_f(\varepsilon_t)\ell_f(\varepsilon_t)'] \in \mathbb{R}^{p \times p}$, where $\ell_f(\varepsilon_t) = (\ell_{f,1}(\varepsilon_t), \dots, \ell_{f,p}(\varepsilon_t))' = -\frac{\dot{f}}{f}(\varepsilon_t)\mathbb{1}[f(\varepsilon_t) > 0]$, is finite.

(c) We have $\mathbb{E}[\varepsilon_t] = 0$, and the covariance matrix $\Sigma = (\sigma_{ij}) = \text{Var}[\varepsilon_t] \in \mathbb{R}^{p \times p}$ is positive definite and finite.

Denote the class of densities satisfying Assumption 3.2.1 by \mathcal{F}_p . Here f is (essentially) unrestricted in the sense that it is unknown and only known to belong to \mathcal{F}_p . Throughout this paper, θ denotes the parameter of interest in the d -dimensional parameter space Θ , while μ and f are nuisance parameters. Our main goal is to develop tests, with attractive features, for the hypothesis

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1 \quad (3.2.2)$$

where $\Theta_0 \cup \Theta_1 = \Theta$.⁵ To proceed with the asymptotic analysis, some preliminary preparations are needed.

Nonparametric modeling for f

To describe local perturbations to the density f , we introduce the separable Hilbert space, with $e = (e_1, \dots, e_p)' \in \mathbb{R}^p$,

$$\mathbf{L}_2^{0,f} = \mathbf{L}_2^{0,f}(\mathbb{R}^p, \mathcal{B}) = \left\{ f_b \in \mathbf{L}_2^f(\mathbb{R}^p, \mathcal{B}) \mid \mathbb{E}f_b(e) = 0, \mathbb{E}ef_b(e) = 0 \right\},$$

where $\mathbf{L}_2^f(\mathbb{R}^p, \mathcal{B})$ denotes the space of Borel-measurable functions $f_b : \mathbb{R}^p \rightarrow \mathbb{R}$ satisfying $\int f_b^2(e)f(e)de < \infty$. Because of the separability, there exists a countable orthonormal basis $b = (b_k)_{k \in \mathbb{N}}$, of $\mathbf{L}_2^{0,f}$ such that $b_k \in C_{2,b}(\mathbb{R})$ for all k (i.e., each b_k is bounded and two times continuously differentiable with bounded derivatives). Hence each function $f_b \in \mathbf{L}_2^{0,f}$ can be written as $f_b = \sum_{k=1}^{\infty} \eta_k b_k$, for some $\eta = (\eta_k)_{k \in \mathbb{N}} \in c_{00}$, where c_{00} is defined as the space of all infinite sequences with only a finite number of non-zero terms. For $b_k \in \mathbf{L}_2^{0,f}$ with $\text{Var } b_k(e) = 1$, $\eta \in c_{00}$, we now introduce the following nonparametric model on f :

$$f_{\eta}^{(T)}(e) = f(e) \left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k(e) \right), \quad (3.2.3)$$

where η is the localized perturbation parameter. Note $f_0^{(T)} = f$. In the following proposition, we show that this way of modeling is valid in the sense

⁵We leave the study about estimation for future work, since the duality of (point) estimation and inference procedure breaks down in nonstandard econometric problems.

that $f_\eta^{(T)}$ satisfies the assumptions on the innovation density that we impose throughout (Assumption 3.2.1). The proof is provided in Appendix 3.7.

Proposition 3.2.1. *Let f satisfy Assumption 3.2.1 and $\eta \in c_{00}$. Then there exists T' such that for all $T \geq T'$, we have $f_\eta^{(T)} \in \mathcal{F}_p$.*

Localization of θ and μ

For $i = 1, \dots, q$ and $j = 1, \dots, p$, let

$$\theta_i = \theta_{0,i} + h_i/T^{\delta_i}, \quad (3.2.4)$$

$$\mu_j = \mu_{0,j} + m_j/\sqrt{T}, \quad (3.2.5)$$

where θ_0 and μ_0 are the (unknown) true values. The convergence rate for θ_i , δ_i , needs to be chosen case-by-case, while the convergence rate for μ_i is $1/2$.⁶ Then, the hypothesis of interest becomes

$$H_0 : h \in \mathcal{H}_0 \quad \text{against} \quad H_1 : h \in \mathcal{H}_1, \quad (3.2.6)$$

where $\mathcal{H}_0 \cup \mathcal{H}_1 = \mathcal{H}$, which three are determined by Θ_0 , Θ_1 , and Θ , respectively. m and η are considered either known or unknown, but nuisance parameters.

Probability measures

Denote by $P_{h,m,\eta;f}^{(T)}$ the law of y_1, \dots, y_T generated by the model (3.2.1), where θ and μ are localized as in (3.2.4) and (3.2.5), and the innovation density is modeled in (3.2.3) with local perturbation parameter η . Also denote the associated probability space by $(\Omega^{(T)}, \mathcal{F}^{(T)}, P_{h,m,\eta;f}^{(T)})$, and the expectation taken under the measure $P_{h,m,\eta;f}^{(T)}$ in this probability space by E . Moreover, let us already mention that we will introduce a collection of probability measures $\mathbb{P}_{h,m,\eta}$ representing the limit experiment in Section 3.2.3. We denote

⁶For example, in unit root testing case, the convergence rate is 1 (see Phillips (1987)).

the probability space associated to the limit experiment by $(\Omega, \mathcal{F}, \mathbb{P}_{h,m,\eta})$, and the expectation taken under $\mathbb{P}_{h,m,\eta}$ by \mathbb{E} .

Partial-sum processes

To describe the limit experiment, it is convenient to introduce some partial sum processes and their limits beforehand. Define, for $u \in [0, 1]$,

$$\begin{aligned} W_\varepsilon^{(T)}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \varepsilon_t, \\ W_{\ell_f}^{(T)}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_f(\varepsilon_t), \\ W_{b_k}^{(T)}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} b_k(\varepsilon_t), \quad k \in \mathbb{N}. \end{aligned} \tag{3.2.7}$$

Using Assumption 3.2.1 we find, under $P_{h,m,0,f}^{(T)}$, weak convergence of $W_\varepsilon^{(T)}$, $W_{\ell_f}^{(T)}$, and $W_{b_k}^{(T)}$ to Brownian motions W_ε , W_{ℓ_f} , and W_{b_k} that are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0,0})$. These Brownian motions belong to $D^p[0, 1]$, $D^p[0, 1]$ and $D[0, 1]$, respectively. Define $W_b(u) = (W_{b_k}(u))' \in D^\infty[0, 1]$, $u \in [0, 1]$. Under H_0 , their covariance matrix is given by

$$\text{Var} \begin{pmatrix} W_\varepsilon(1) \\ W_{\ell_f}(1) \\ W_b(1) \end{pmatrix} = \begin{pmatrix} \Sigma & I_p & 0_{p,\infty} \\ I_p & J_f & J_{fb} \\ 0'_{p,\infty} & J_{bf} & I_\infty \end{pmatrix}, \tag{3.2.8}$$

where $J_{fb} = (J_{fb}[i, j]) \in \mathbb{R}^{p \times K} = \mathbb{E}[\ell_f(\varepsilon_t)b(\varepsilon_t)']$ and $J_{bf} = J_{fb}'$, Σ and J_f are defined in Assumption 3.2.1, I_p and I_∞ are identity matrices of dimension p and ∞ respectively, and $0_{p,\infty}$ is a $p \times \infty$ dimensional zero matrix. Note that e and $b_k(e)$ form an orthogonal basis of \mathbb{L}_2^f . Therefore, all one-dimensional Brownian motions can be decomposed into a combination of W_ε and W_{b_k} , $k \in \mathbb{N}$. This establishes the following decomposition

$$W_{\ell_f} = \Sigma^{-1}W_\varepsilon + J_{fb}W_b, \quad \text{and} \quad J_f = \Sigma^{-1} + J_{fb}J_{bf}. \tag{3.2.9}$$

Limits of nonrandom variables x_t and z_t

Likewise, for convenience's sake, we also introduce limit representations of the nonrandom explanatory variables. First, we impose the following assumption.

Assumption 3.2.2. For some positive definite $k \times k$ matrix D_{zz} , $T^{-1} \sum_{t=1}^T z_t z_t' \rightarrow D_{zz}$ and $\max_{1 \leq t \leq T} \|z_t\|/\sqrt{T} \rightarrow 0$. For some positive definite $q \times q$ matrix D_{xx} , $T^{-1} \sum_{t=1}^n x_t x_t' \rightarrow D_{xx}$ and $\max_{1 \leq t \leq T} \|x_t\|/\sqrt{T} \rightarrow 0$. $T^{-1} \sum_{t=1}^n z_t x_t' \rightarrow 0$.

Define, for $u \in [0, 1]$, $x^{(T)}(u) = x_{\lfloor uT \rfloor}$ and $z^{(T)}(u) = z_{\lfloor uT \rfloor}$. As $T \rightarrow \infty$, they respectively converge to two deterministic processes x_u and z_u . By Assumption 3.2.2, x_u and z_u are square integrable with $D_{xx} = \int_0^1 x_u x_u' du$ and $D_{zz} = \int_0^1 z_u z_u' du$. Also, we define $\mu_x = \int_0^1 x_u du$ and $\mu_z = \int_0^1 z_u du$ as the means of these two deterministic processes.

Running Example (Linear Regression Model). Let $X_T = (x_1, \dots, x_T)' \in \mathbb{R}^{T \times q}$ be observations of a p -dimensional exogenous nonrandom variable, and $Y_T = (y_1, \dots, y_T)' \in \mathbb{R}^{T \times 1}$ be observations of the dependent one-dimensional variable. The linear regression model is specified by

$$y_t = \theta' x_t + \mu + \varepsilon_t, \quad t = 1, \dots, T,$$

where $(\varepsilon_1, \dots, \varepsilon_T)' \in \mathbb{R}^{T \times 1}$ are i.i.d. innovations with density $f \in \mathcal{F}_1$. Localize the parameters as $\theta = \theta_0 + h/\sqrt{T}$ and $\mu = \mu_0 + m/\sqrt{T}$. We are interested in testing the null hypothesis $H_0 : h = 0$ against the alternative hypothesis $H_1 : h \neq 0$ (so that $p = 1$, $d = q$, $\mathcal{H}_0 = \{0\}$, and $\mathcal{H}_1 = \mathbb{R}^d \setminus \mathcal{H}_0$). The density f is nonparametrically modeled as in (3.2.3). We define the partial sum processes, for $u \in [0, 1]$, $W_\varepsilon^{(T)}(u) = T^{-1} \sum_{t=1}^{\lfloor uT \rfloor} \varepsilon_t$, $W_{\ell_f}^{(T)}(u) = T^{-1} \sum_{t=1}^{\lfloor uT \rfloor} \ell_f(\varepsilon_t)$ and $W_{b_k}^{(T)}(u) = T^{-1} \sum_{t=1}^{\lfloor uT \rfloor} b_k(\varepsilon_t)$, where $\varepsilon_t = y_t - \mu_0$ and $\ell_f = -\dot{f}/f$. Under H_0 , $W_\varepsilon^{(T)}(u) \Rightarrow W_\varepsilon(u)$, $W_{\ell_f}^{(T)}(u) \Rightarrow W_{\ell_f}(u)$ and $W_{b_k}^{(T)}(u) \Rightarrow W_{b_k}(u)$.

The covariance matrix is as in equation (4.2.15), with $p = 1$ and Σ being a constant. \square

3.2.2 Limit Likelihood Ratios

Denote the log-likelihood ratio $\log(d\mathbb{P}_{h,m,\eta}^{(T)}/d\mathbb{P}_{0,0,0}^{(T)})$ by $L^{(T)}(h, m, \eta)$. Under mild assumptions, it weakly converges to a limit log-likelihood ratio denoted by $\mathcal{L}(h, m, \eta)$. For many cases, $\mathcal{L}(h, m, \eta)$ is of the following form

$$\mathcal{L}(h, m, \eta) = \Delta(h, m, \eta) - \frac{1}{2}\mathcal{Q}(h, m, \eta), \quad (3.2.10)$$

with

$$\begin{aligned} \Delta(h, m, \eta) &= \int_0^1 [M(u; h) + m]' dW_{\ell_f}(u) + \eta' dW_b(1), \\ \mathcal{Q}(h, m, \eta) &= \int_0^1 [M(u; h) + m]' J_f [M(u; h) + m] du \\ &\quad + 2 \int_0^1 [M(u; h) + m]' J_{fb} \eta du + \eta' \eta, \end{aligned}$$

where $M(u; h)$ is a stochastic process specified by the given function ϕ in (3.2.1). In Jeganathan (1995), this type of limit experiment is called *Locally Asymptotically Brownian Functional* (LABF).

Normally we have, for any value of $h, m \in \mathbb{R}$ and $\eta \in c_{00}$, the expectation of $\mathcal{L}(h, m, \eta)$ is one under $\mathbb{P}_{0,0,0}$. This allows us to introduce a new collection of probability measures, $\mathbb{P}_{h,m,\eta}$, on the measurable space (Ω, \mathcal{F}) by their Radon-Nikodym derivatives with respect to $\mathbb{P}_{0,0,0}$:

$$\frac{d\mathbb{P}_{h,m,\eta}}{d\mathbb{P}_{0,0,0}} = \exp(\mathcal{L}(h, m, \eta)).$$

Moreover, as $T \rightarrow \infty$, the sequence of experiences (indexed by T) introduced in the beginning of Section 3.2.1 weakly converges (in the Le Cam sense) to a limit experiment described by the probability measures $\mathbb{P}_{h,m,\eta}$. Denote this limit experiment by $\mathcal{E}(f)$, in Section 3.2.3 below, we will provide a structural version of it.

Running Example—Continued (Linear Regression Model). The limit experiment of the linear regression model, e.g., Van der Vaart (2000), is LAN. Specifically, the log-likelihood ratio of the sequence of experiments is

$$L^{(T)}(h, m, \eta) = \Delta^{(T)}(h, m, \eta) - \frac{1}{2}\mathcal{Q}(h, m, \eta) + o_{p_0}(1) \quad (3.2.11)$$

where

$$\begin{aligned} \Delta^{(T)}(h, m, \eta) &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (h'x_t + m)\ell_f(\varepsilon_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta'b(\varepsilon_t), \\ \mathcal{Q}(h, m, \eta) &= J_f h' D_{xx} h + m^2 J_f + \eta'\eta + 2h'\mu_x J_{fb}\eta + 2h'\mu_x J_f m + 2m J_{fb}\eta. \end{aligned}$$

where $\varepsilon_t = y_t - \mu_0$. Then we have, under $P_{0,0,0}^{(T)}$, $\Delta^{(T)}(h, m, \eta)$ weakly converges to $\Delta(h, m, \eta) = \int_0^1 (h'x_u + m) dW_{\ell_f}(u) + \eta'W_b(1)$, which is normally distributed with mean zero and variance $Q(h, m, \eta)$. This is due to the fact that x_u is a deterministic process in the limit. Define the limit likelihood ratio $\mathcal{L}(\vartheta, m, \eta) = \Delta(h, m, \eta) - \frac{1}{2}\mathcal{Q}(h, m, \eta)$. \square

3.2.3 Structural Limit Experiment

In the literature, the limit experiment derivation often stops at achieving the limit likelihood ratio, while in the present paper, we show that a structural representation of the limit experiment can be given and exploited. A direct application of Girsanov's Theorem on the limit likelihood ratio in equation (3.2.10) gives the structural limit experiment in the following theorem.

Theorem 3.2.1 (Structural Limit Experiment $\mathcal{E}(f)$). *Suppose we observe processes $W_\varepsilon \in D^p[0, 1]$, $W_{\ell_f} \in D^p[0, 1]$, and $W_b \in D^\infty[0, 1]$ modeled as*

$$dW_\varepsilon(u) = M(u, h)du + mdu + dZ_\varepsilon(u), \quad (3.2.12)$$

$$dW_b(u) = J_{bf}M(u, h)du + J_{bf}mdu + \eta du + dZ_b(u), \quad (3.2.13)$$

$$dW_{\ell_f}(u) = J_f M(u, h)du + J_f mdu + J_{fb}\eta du + dZ_{\ell_f}(u), \quad (3.2.14)$$

where Z_ε , Z_{ℓ_f} and Z_b are multivariate Brownian motions with covariance matrix defined in (4.2.15), then the log-likelihood ratio $\log(d\mathbb{P}_{h,m,\eta}/d\mathbb{P}_{0,0,0})$ equals that in (3.2.10).

Actually, (3.2.14) can be omitted since it could be implied by (3.2.12) and (3.2.13), due to the decomposition relation in (3.2.9). In this limit experiment $\mathcal{E}(f)$, the hypothesis of interest is still

$$H_0 : h \in \mathcal{H}_0 \quad \text{against} \quad H_1 : h \in \mathcal{H}_1, \quad (3.2.15)$$

and m and η are nuisance parameters. Thus, our goal is to first conduct inference of desirable properties in $\mathcal{E}(f)$, and then find some statistics in the original sequence of experiments that can behave similarly as $T \rightarrow \infty$.

Running Example—Continued (Linear Regression Model). The structural limit experiment for the linear regression model, implied by the limit likelihood ratio $L(h, \eta)$, is

$$\begin{aligned} dW_\varepsilon(u) &= h'x_u du + mdu + dZ_\varepsilon(u), \\ dW_b(u) &= J_{bf}h'x_u du + J_{bf}mdu + \eta du + dZ_b(u), \\ dW_{\ell_f}(u) &= J_f h'x_u du + J_f mdu + J_{fb}\eta du + dZ_{\ell_f}(u), \end{aligned}$$

where Z_ε , Z_{ℓ_f} , and Z_b are multivariate Brownian motions with covariance matrix (4.2.15). \square

3.2.4 Maximal Invariant and Semiparametric Efficiency Bounds

In the structural limit experiment derived in Theorem 3.2.1, the parameter of interest is h , while m and η are regarded as nuisance parameters. Since it only shows up in the drift part of the model, we suggest to eliminate η by imposing the invariant constraint, to be specific, by taking Brownian bridges. Introduce the transformation $\bar{\mathbf{g}}_{0,\eta} : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ on η by

$$\bar{\mathbf{g}}_{0,\eta}\eta = \eta + c_\eta,$$

and the transformation $\bar{\mathfrak{g}}_{m,0} : \mathbb{R}^p \rightarrow \mathbb{R}^p$ on m by

$$\bar{\mathfrak{g}}_{m,0}m = m + c_m.$$

where $c_\eta \in c_{00}$ and $c_m \in \mathbb{R}^p$ are some vectors. Define $\bar{\mathfrak{g}}_{m,\eta} = (\bar{\mathfrak{g}}_{m,0}; \bar{\mathfrak{g}}_{0,\eta})$. Also, denote by $\mathfrak{g}_{0,\eta}$, $\mathfrak{g}_{m,0}$ and $\mathfrak{g}_{m,\eta}$ the sample-space counterparts of the above parameter-space transformations, and by $\mathfrak{G}_{0,\eta}$, $\mathfrak{G}_{m,0}$ and $\mathfrak{G}_{m,\eta}$ their belonging groups. Intuitively, in the structural limit experiment $\mathcal{E}(f)$, the transformation $\mathfrak{g}_{0,\eta}$ adds a drift $u \mapsto c_\eta du$ to the process $W_b(u)$; the transformation $\mathfrak{g}_{m,0}$ adds drifts $u \mapsto c_m du$ and $u \mapsto J_{bf}c_m du$ respectively to the processes $W_\varepsilon(u)$ and $W_b(u)$; and the transformation $\mathfrak{g}_{m,\eta}$ adds drifts $u \mapsto c_m du$ and $u \mapsto (J_{bf}c_m + c_\eta)du$ respectively to the processes $W_\varepsilon(u)$ and $W_b(u)$.

A natural invariant statistic with respect to a drift transformation on a process is the bridge of that process. To be specific, for $u \in [0, 1]$, let $B^{W(u)} := W(u) - uW(1)$, which is intuitively “taking the bridge of the process $W(u)$ ”. We have, $B^{W(u)+cu} = W(u) + cu - u(W(1) + c) = W(u) - uW(1) = B^{W(u)}$, where c is an arbitrary constant vector with the same dimension as $W(u)$. Another type of statistic, which is invariant with respect to a transformation which adds drifts of certain proportion respectively to some processes, is by taking a certain combination of these processes based on that proportion (see $\mathcal{M}_{m,0}$ in (3.2.17) as an example). Next, the discussion continues with three different cases regarding the knowledge of m and η .

When $m = 0$ is known and η is unknown, a candidate invariant statistic is

$$\mathcal{M}_{0,\eta} = \sigma(W_\varepsilon(u), B_b(u), u \in [0, 1]); \quad (3.2.16)$$

and when $\eta = 0$ is known and m is unknown, a candidate invariant statistic is

$$\mathcal{M}_{m,0} = \sigma(J_{bf}W_\varepsilon(u) - W_b(u), B_b(u), u \in [0, 1]); \quad (3.2.17)$$

and when both m and η are unknown, a candidate invariant statistic is

$$\mathcal{M}_{m,\eta} = \sigma(B_\varepsilon(u), B_b(u), u \in [0, 1]). \quad (3.2.18)$$

Actually, in the following theorem, we show that these candidate invariant statistics are the maximal invariant statistics in the corresponding cases. Its proof is again provided in Appendix 3.7.

Theorem 3.2.2 (Maximal Invariant). *In the structural limit experiment $\mathcal{E}(f)$:*

- (a) $\mathcal{M}_{0,\eta}$ in (3.2.16) is maximal invariant with respect to $\mathfrak{G}_{0,\eta}$,
- (b) $\mathcal{M}_{m,0}$ in (3.2.17) is maximal invariant with respect to $\mathfrak{G}_{m,0}$,
- (c) $\mathcal{M}_{m,\eta}$ in (3.2.18) is maximal invariant with respect to $\mathfrak{G}_{m,\eta}$.

Theorem 3.2.2 implies that any inference which is invariant with respect to $\mathfrak{G}_{i,j}$ must be a function of $\mathcal{M}_{i,j}$, where $\{i, j\} \in \mathcal{S} := \{\{0, \eta\}, \{m, 0\}, \{m, \eta\}\}$. Thus, by the Neyman-Pearson lemma, the power function of the likelihood ratio test based on the observation $\mathcal{M}_{i,j}$ provides an upper bound for invariant tests. The log-likelihood ratio based on $\mathcal{M}_{i,j}$, denoted by $\mathcal{L}_{\mathcal{M}_{i,j}}$, can be calculated using the following conditional expectation formula

$$\mathcal{L}_{\mathcal{M}_{i,j}}(h) = \log E_0 [\exp(\mathcal{L}(h, i, j)) | \mathcal{M}_{i,j}], \quad \{i, j\} \in \mathcal{S}. \quad (3.2.19)$$

Define the associated likelihood ratio test $\phi(h, \alpha) := \mathbb{1}\{\mathcal{L}_{\mathcal{M}_{i,j}}(h) > \kappa_\alpha(h)\}$, where $\kappa_\alpha(h)$ is the $1 - \alpha$ quantile of $\mathcal{L}_{\mathcal{M}_{i,j}}(h)$. Then, the (semiparametric) power envelop of all asymptotically invariant tests is given by

$$\Psi(h, \alpha) = E_0 \left[\phi(h, \alpha) \frac{d\mathbb{P}_{h,m,\eta}}{d\mathbb{P}_{0,m,\eta}} \right]. \quad (3.2.20)$$

Remark 3.2.1 (Sufficient Statistic). Note that, when the nuisance parameter η is unknown, the corresponding part in the limit central sequence $\Delta(h, m, \eta)$, $\eta'W_b(1)$, is always independent of the maximal invariant $\mathcal{M}_{0,\eta}$ or $\mathcal{M}_{m,\eta}$. As a result, η vanishes as well as W_b after taking the conditional expectation on these maximal invariant statistics. Therefore, with the knowledge of Σ , J_f and $M(u, h)$ in the limit, $\sigma(W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$ becomes a sufficient statistic for h . Henceforth, for the sake of convenience, we omit the equation for W_b (like equation (3.2.13) in Theorem 3.2.1 in the structural limit experiment $\mathcal{E}(f)$, and replace B_b in the maximal invariant statistics by B_{ℓ_f} . Specifically, define the “simplified” version of the structural limit experiment, namely $\mathcal{E}^*(f)$,

$$dW_\varepsilon(u) = M(u, h)du + mdu + dZ_\varepsilon(u), \quad (3.2.21)$$

$$dW_{\ell_f}(u) = J_f M(u, h)du + J_f mdu + J_{fb}\eta du + dZ_{\ell_f}(u), \quad (3.2.22)$$

and the “simplified” maximal invariant statistics $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$, $\mathcal{M}_{m,0}^* = (J_f W_\varepsilon(u) - W_{\ell_f}(u), B_{\ell_f}(u), u \in [0, 1])$ and $\mathcal{M}_{m,\eta}^* = (B_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$, with everything else left the same as in $\mathcal{E}(f)$. Notably, a similar proof as that for Theorem 3.2.2 can show that $\mathcal{M}_{i,j}^*$ are indeed the maximal invariant statistics with respect to the group of transformations $\bar{\mathbf{g}}_{i,j}$ in $\mathcal{E}^*(f)$, $i, j \in \mathcal{S}$. Once again, this step is only for simplicity and causes no information loss, i.e., $E \left[d\mathbb{P}_{h,i,j} / d\mathbb{P}_{0,0,0} | \mathcal{M}_{i,j}^* \right] = E \left[d\mathbb{P}_{h,i,j} / d\mathbb{P}_{0,0,0} | \mathcal{M}_{i,j} \right]$ (or $\mathcal{L}_{\mathcal{M}_{i,j}^*}(\theta) = \mathcal{L}_{\mathcal{M}_{i,j}}(\theta)$). Then, in the rest of this paper, we focus on the simplified structural limit experiment $\mathcal{E}^*(f)$.

Running Example—Continued (Linear Regression Model). In the Linear Regression model, assume the location parameter μ (or m) for the innovation distribution is unknown.⁷ A rewrite of the simplified structural limit experiment $\mathcal{E}^*(f)$ is

$$W_\varepsilon(u) = hx_u du + mdu + Z_\varepsilon(u),$$

⁷It is the same to assume the first component of x_t is 1, and the associated local parameter h_1 is unknown and of no interest, since m and h_1 are not identified.

$$W_{\ell_f}(u) = J_f h x_u du + J_f m du + J_{fb} \eta du + Z_{\ell_f}(u),$$

By Theorem 3.2.2, the maximal invariant statistic is $\mathcal{M}_{m,\eta}^* = (B_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$. The limit log-likelihood ratio of $\mathcal{M}_{m,\eta}^*$ is

$$\mathcal{L}_{\mathcal{M}_{m,\eta}^*}(h) = \Delta_{\mathcal{M}_{m,\eta}^*}(h) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{m,\eta}^*}(h), \quad (3.2.23)$$

where

$$\begin{aligned} \Delta_{\mathcal{M}_{m,\eta}^*}(h) &= \int_0^1 h' x_u dB_{\ell_f}(u), \\ \mathcal{Q}_{\mathcal{M}_{m,\eta}^*}(h) &= J_f [h' D_{xx} h - (h' \mu_x)^2]. \end{aligned}$$

Note that $\Delta_{\mathcal{M}_{m,\eta}^*}(h) = h' \int_0^1 x_u dB_{\ell_f}(u) = h' \int_0^1 (x_u - \mu_x) dW_{\ell_f}(u)$, which is exactly the limit of the traditional efficient score $\frac{1}{\sqrt{T}} \sum_{t=1}^T h' (x_t - \bar{x}) \ell_f(\varepsilon_t)$ with $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ (see Bickel et al. (1998)).

It is also worth noting that in the case where m is unknown and of no interest while $\eta = 0$ is known, the maximal invariant, according to Theorem 3.2.2 and Remark 3.2.1, is $\mathcal{M}_{m,0}^* = \sigma(J_f W_\varepsilon(u) - W_{\ell_f}(u), B_{\ell_f}(u), u \in [0, 1])$. The limit log-likelihood ratio of $\mathcal{M}_{m,0}^*$ is

$$\mathcal{L}_{\mathcal{M}_{m,0}^*}(h) = \Delta_{\mathcal{M}_{m,0}^*}(h) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{m,0}^*}(h), \quad (3.2.24)$$

where

$$\begin{aligned} \Delta_{\mathcal{M}_{m,0}^*}(h) &= \int_0^1 h' x_u dB_{\ell_f}(u), \\ \mathcal{Q}_{\mathcal{M}_{m,0}^*}(h) &= J_f [h' D_{xx} h - (h' \mu_x)^2], \end{aligned}$$

which is exactly the same as $\mathcal{L}_{\mathcal{M}_{m,\eta}^*}(h)$. This refers to the concept of adaptivity: the inference for the parameter of interest h can work equally well with or without the knowledge of the nonparametric part η . Note that here this new approach leads to the same conclusion as that of (Newey, 1990, Section 3) for the same linear regression model. \square

3.2.5 Semiparametric Efficient Inference with Rank Statistics

The appearance of the Brownian bridges (under H_0) in the previous section is suggestive: they are invariant with respect to transformations on both the

location parameter m and the density disturbance parameter η in the limit, while statistics based on ranks and a reference density enjoy distribution-freeness property in the sequence, and they converge weakly to Brownian bridges in the limit.

In this section, let us restrict to the simpler one-dimensional innovation case ($p = 1$). Under the null hypothesis, we have that $\varepsilon_t = y_t - \phi(y_{t-1}, x_t, z_t; \theta_0) - \mu_0$. Denote by R_t the rank of ε_t , $t = 1, \dots, T$. Since f is unknown, we let $g \in \mathcal{F}_1$ be a given (so-called *reference*) density, and denote by $\ell_g = -\dot{g}/g$ its score function and by G^{-1} its the inverse cumulative density function. Define the partial sum process by

$$W_{\ell_g}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_g(\varepsilon_t), \quad (3.2.25)$$

$$B_{\ell_g}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor uT \rfloor} \ell_g \left[G^{-1} \left(\frac{R_t}{T+1} \right) \right]. \quad (3.2.26)$$

Then, by Lemma A.1 in Hallin, Van den Akker, and Werker (2011), we have

$$W_{\ell_g}^{(T)}(u) \Rightarrow W_{\ell_g}(u) \quad \text{and} \quad B_{\ell_g}^{(T)}(u) \Rightarrow B_{\ell_g}(u), \quad (3.2.27)$$

where W_{ℓ_g} denotes a zero-drift Brownian motion with variance J_g per unit of time and B_{ℓ_g} its associated Brownian bridge: $B_{\ell_g}(u) = W_{\ell_g}(u) - uW_{\ell_g}(1)$, $u \in [0, 1]$. The limit behavior of $B_{\ell_g}(u)$ is described by the covariance matrix

$$\text{Var} \begin{pmatrix} W_{\varepsilon}(1) \\ W_{\ell_f}(1) \\ W_{\ell_g}(1) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 1 & \sigma_{\varepsilon g} \\ 1 & J_f & J_{fg} \\ \sigma_{\varepsilon g} & J_{fg} & J_g \end{pmatrix},$$

with

$$\begin{aligned} \sigma_{\varepsilon g} &= \int_0^1 F^{-1}(u) \ell_g(G^{-1}(u)) \, du, \\ J_{fg} &= \int_0^1 \ell_f(F^{-1}(u)) \ell_g(G^{-1}(u)) \, du. \end{aligned}$$

So then, the limit of references based on $W_{\varepsilon}^{(T)}$ in (3.2.7) and $B_{\ell_g}^{(T)}$ in (3.2.26) can be expressed with W_{ε} , B_{ℓ_g} , $\sigma_{\varepsilon g}$, J_g and J_{fg} . When $g = f$, it holds that $B_{\ell_g} = B_{\ell_f}$.

However, for rank statistics, it is not easy to extend from the univariate case to multivariate case. This is because vectors cannot be ordered without imposing any additional measure. In Section 3.3, I will propose a new way to do this.

Running Example—Continued (Linear Regression Model). The semiparametric inference based on rank statistics for the linear regression model is

$$\hat{L}_g^{(T)}(h) = \hat{\Delta}_g^{(T)}(h) - \frac{1}{2}\hat{Q}_g^{(T)}(h),$$

where

$$\hat{\Delta}_g^{(T)}(h) = \int_0^1 h' x_u^{(T)} dB_{\ell_g}^{(T)}(u), \quad \hat{Q}_g^{(T)}(h) = J_g [h' D_{xx} h + (h' \mu_x)^2].$$

3.2.6 Some Discussions

Remark 3.2.2 (Exact scores). The partial sum process $B_{\ell_g}^{(T)}$ in (3.2.26) is actually based on the so-called “approximate score” in the rank statistic literature. Another version based on what called the “exact score” is given by

$$B_{\ell_g}^{*(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor uT \rfloor} E[\ell_g(G(\varepsilon_t)) | R_t = \lfloor u(T+1) \rfloor], \quad u \in (0, 1).$$

The exact score version is more convenient for proofs since $B_{\ell_g}^{*(T)}(1)$ identically equals zero, while the approximate score version is more convenient to implement. They share the same limit as $T \rightarrow \infty$, and as for finite sample performance, they behave similarly. Therefore, we focus on the approximate version throughout this paper.

Remark 3.2.3 (Serial-correlated innovations). For serial-correlated innovations, we can adjust the rank-based inference introduced in Section 3.2.5 for the univariate cases, or the one will be introduced below in Section 3.3 for the multivariate cases, by using the “aligned” rank statistics. More specifically, we regress $\hat{\varepsilon}_t$ on its lagged values $\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots$, and use the ranks of

the obtained residuals. Since the inference for the parameter of interest is adaptive with respect to these autocorrelation coefficients in many models, as a consequence, the asymptotic results of the i.i.d. case above remain unchanged.

Remark 3.2.4 (LAN). In the LAN type limit experiments (also known as Gaussian shift experiments), the observation is a single draw from a normal distribution. This single observation is exactly $\int_0^1 M(u, h) W_{\ell_f}(u)$ with $M(u, h)$ being deterministic, which henceforth is normally distributed with mean zero and variance $J_f \int_0^1 M(u, h)^2 du$. This provides an alternative perspective for the traditional LAN models. Moreover, this structural limit experiment would suggest test statistics in nonstandard problems, and we will use this for our second application about weak instruments in the present paper.

Remark 3.2.5 (Adaptivity). To see if the restriction to invariant tests is without power loss (the adaptivity property) using the new approach in the present paper, one can simply compare the associated likelihood ratio of the maximal invariant with the likelihood ratio obtained by using the knowledge of the associated nuisance parameter. In the linear regression example of this section, we have shown above that the inference for h is adaptive with respect to the nuisance parameter η when m is unknown. This adaptivity result also applies to the weak instrument example in Section 3.5. Nevertheless, this adaptivity property breaks in the cointegration case (see Section 3.4) as we lose some efficiency by imposing the invariance restriction to eliminate η .

3.3 Multivariate Rank Statistic

The extension of rank statistics from the univariate case to the multivariate case is not straightforward, since \mathbb{R}^p , $p \geq 2$, is not naturally ordered. To avoid this, in this paper, we propose methods of constructing test statistics based on component-wise ranks.

3.3.1 Standardized Error ν_t

Suppose ε_t follows a distribution with density $f \in \mathcal{F}_p$. Assuming the covariance matrix Σ is known, we introduce the standardized error defined as

$$\nu_t = \Sigma^{-\frac{1}{2}} \varepsilon_t.$$

Accordingly, the density of ν_t is $f_\nu(\nu_t) = |\Sigma|^{-\frac{1}{2}} f(\Sigma^{\frac{1}{2}} \nu_t)$. We call ν_t the standardized error in the sense that its p elements are uncorrelated, i.e., the covariance matrix of ν_t is the p -dimensional identity matrix I_p . The score function of f_ν is

$$\ell_{f_\nu}(\nu_t) = -\frac{\dot{f}_\nu}{f_\nu}(\nu_t) = -\Sigma^{\frac{1}{2}} \frac{\dot{f}}{f}(\Sigma^{\frac{1}{2}} \nu_t) = \Sigma^{\frac{1}{2}} \ell_f(\Sigma^{\frac{1}{2}} \nu_t),$$

and, subsequently, the Fisher information matrix of f_ν is ΣJ_f .

Define the following two partial sum processes, for $u \in [0, 1]$, as

$$\begin{aligned} W_\nu^{(T)}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \nu_t, \\ W_{\ell_{f_\nu}}^{(T)}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_{f_\nu}(\nu_t). \end{aligned}$$

By Functional Central Limit Theorem, they weakly converge to W_ν and $W_{\ell_{f_\nu}}$, respectively. The covariance matrix of them at time unity is given by

$$\text{Var} \begin{pmatrix} W_\nu(1) \\ W_{\ell_{f_\nu}}(1) \end{pmatrix} = \begin{pmatrix} I_p & I_p \\ I_p & J_{f_\nu} \end{pmatrix}.$$

Followed from the construction above, we immediately have the relationships

$$W_\nu = \Sigma^{-\frac{1}{2}} W_\varepsilon, \quad (3.3.1)$$

$$W_{\ell_{f\nu}} = \Sigma^{\frac{1}{2}} W_{\ell_f}, \quad (3.3.2)$$

$$J_{f\nu} = \Sigma^{\frac{1}{2}} J_f \Sigma^{\frac{1}{2}'} . \quad (3.3.3)$$

Moreover, by Kagana and Landsmanb (1999), $J_{f\nu}$ is a diagonal matrix.

3.3.2 Rank-based Score and Its Limit Behaviour

Similarly as the univariate case in Section 3.2.5, we introduce the *marginal reference density* function $g_i(\cdot)$ for $i = 1, \dots, p$, which can be freely chosen by researchers. Denote the associated cumulative distribution function (CDF) by $G_i(\cdot)$ and the quantile function (or inverse CDF) by $G_i^{-1}(\cdot)$. The following assumption is made on these marginal reference densities.

Assumption 3.3.1. The marginal reference densities g_i , for $i = 1, \dots, p$, are positive, of variance unity, absolutely continuous with a.e. existing derivative \dot{g}_i . Moreover, they satisfy

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left(-\frac{\dot{g}_i}{g_i} \left[G_i^{-1} \left(\frac{t}{T+1} \right) \right] \right)^2 = J_{g_i},$$

where J_{g_i} is the Fisher information for location associated to $g_i(\cdot)$.

Let $R_{i,t}$ be the rank of $\nu_{i,t}$ among $\{\nu_{i,1}, \dots, \nu_{i,T}\}$, and $R_t = (R_{1,t} \cdots R_{p,t})'$ be the associated rank vector of ν_t . Define the reference score function based on rank vector R_t by

$$\ell_g(R_t) := \left(\ell_{g_1} \left[G_1^{-1} \left(\frac{R_{1,t}}{T+1} \right) \right], \dots, \ell_{g_p} \left[G_p^{-1} \left(\frac{R_{p,t}}{T+1} \right) \right] \right)',$$

where $\ell_{g_i}(\cdot) := -\dot{g}_i/g_i(\cdot)$ is the marginal score function. Now we introduce the rank-based score partial sum process as

$$B_{\ell_g}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_g(R_t),$$

whose limit behavior is given by the following theorem. The associated proof is provided in Appendix 3.7.

Theorem 3.3.1. *Let the error term ν_t , $t = 1, \dots, T$, be i.i.d. with density $f_\nu \in \mathcal{F}_p$, and R_t the corresponding rank vector. Suppose the chosen marginal reference densities g_i satisfying Assumption 3.3.1, for $i = 1, \dots, p$, are of unit variance. Then we have*

- (a) $B_{\ell_g}^{(T)} \Rightarrow B_{\ell_g}$, where B_{ℓ_g} is a Brownian bridge defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The convergence is on $D^p[0, 1]$ equipped with the uniform topology.
- (b) Denote by W_{ℓ_g} the associated Brownian motion of B_{ℓ_g} , by $F_{\nu,i}^{-1}$ the quantile function of the marginal distribution $F_{\nu,i}$ of $\nu_{i,t}$, and by $F_{\nu,ij}$ the joint distribution of $\nu_{i,t}$ and $\nu_{j,t}$. The covariance matrix of W_ν , $W_{\ell_{f_\nu}}$ and W_{ℓ_g} is thus given by

$$\text{Var} \begin{pmatrix} W_\nu(1) \\ W_{\ell_{f_\nu}}(1) \\ W_{\ell_g}(1) \end{pmatrix} = \begin{pmatrix} I_p & I_p & \Sigma_{\nu g} \\ I_p & J_{f_\nu} & J_{f_\nu g} \\ \Sigma_{g\nu} & J_{f_\nu g} & J_g \end{pmatrix} \quad (3.3.4)$$

where, for $i, j = 1, \dots, p$,

$$\begin{aligned} \Sigma_{\nu g}[i, j] &= \Sigma_{g\nu}[j, i] = \int_0^1 \int_0^1 F_{\nu,i}^{-1}(F_{\nu,i}(e_i)) \ell_{g_j} \left[G_j^{-1}(F_{\nu,j}(e_j)) \right] dF_{\nu,ij}(e_i, e_j), \\ J_g[i, j] &= \int_0^1 \int_0^1 \ell_{g_i} \left[G_i^{-1}(F_{\nu,i}(e_i)) \right] \ell_{g_j} \left[G_j^{-1}(F_{\nu,j}(e_j)) \right] dF_{\nu,ij}(e_i, e_j), \\ J_{f_\nu g} &= \text{diag}(J_{f_\nu g,1}, \dots, J_{f_\nu g,p}) \quad \text{with} \quad J_{f_\nu g,i} = \int_0^1 \ell_{f_\nu,i} \left[F_{\nu,i}^{-1}(u) \right] \ell_{g_i} \left[G_i^{-1}(u) \right] du. \end{aligned}$$

- (c) $J_{f_\nu g,i} \geq 1$ when g_i is Gaussian, and the equality holds if $f_{\nu,i}$ is also Gaussian.

However, in practice, the covariance matrix Σ is unknown and needed to be estimated. This makes the problem a bit more complicated since we

are actually using the ranks of some estimated standardized error ν_t rather than R_t (or, the estimated ranks), which relates to the traditional *aligned ranks* problem.

Define $\hat{\nu}_t := \hat{\Sigma}^{-\frac{1}{2}} \varepsilon_t$ where $\hat{\Sigma}$ is some estimator of Σ . Let $\hat{R}_{i,t}$ be the rank of $\hat{\nu}_{i,t}$ among $\{\hat{\nu}_{i,1}, \dots, \hat{\nu}_{i,T}\}$, and $\hat{R}_t = (\hat{R}_{1,t} \cdots \hat{R}_{p,T})'$ be the rank vector of $\hat{\nu}_t$. Define the partial-sum processes

$$\widehat{W}_\nu^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \hat{\nu}_t, \quad (3.3.5)$$

$$\widehat{B}_{\ell_g}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_g(\hat{R}_t), \quad (3.3.6)$$

which we will use in practice to construct rank-based test statistics. Moreover, we assume the following assumption on the estimator of covariance matrix $\hat{\Sigma}$.

Assumption 3.3.2. There exists consistent, under the null hypothesis, estimator $\hat{\Sigma}$ of Σ such that

$$\widehat{W}_\nu^{(T)} \Rightarrow W_\nu \quad \text{and} \quad \widehat{B}_{\ell_g}^{(T)} \Rightarrow B_{\ell_g}. \quad (3.3.7)$$

This assumption indicates that, with well-chosen consistent covariance matrix estimator $\hat{\Sigma}$, the feasible partial-sum processes $\widehat{W}_\nu^{(T)}$ and $\widehat{B}_{\ell_g}^{(T)}$ have the same limit behaviors as the partial-sum processes $W_\nu^{(T)}$ and $B_{\ell_g}^{(T)}$, respectively. Moreover, we conjecture that the consistency of $\hat{\Sigma}$ would automatically lead to the convergence result in (3.3.7). The proof of this conjecture relates to the *align rank* problem. We leave the details for future work.

3.3.3 Construction of Rank-based Statistical Inference

Till now, we have introduced the feasible partial sum process of the residuals $\widehat{W}_\nu^{(T)}$, and the rank-based partial sum process of the reference scores $\widehat{B}_{\ell_g}^{(T)}$.

For the purpose of constructing efficient statistical inference, we come up with the following two methods based on these partial sum processes.

Method I: inference procedures based on $\sigma(W_\nu, B_{\ell_g})$

The first method is to construct inference based on the likelihood ratio of the σ -field of W_ν and B_{ℓ_g} , defined by $\mathcal{M}_g := \sigma(W_\nu, B_{\ell_g})$. This likelihood ratio can be derived through the following conditional expectation

$$\begin{aligned}\mathcal{L}_{\mathcal{M}_g}(h; W_\nu, B_{\ell_g}) &= \log \mathbb{E} [\mathcal{L}_{\mathcal{M}_{i,j}}(h) | \sigma(W_\nu, B_{\ell_g})], \\ &= \Delta_{\mathcal{M}_g}(h; W_\nu, B_{\ell_g}) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_g}(h; W_\nu, B_{\ell_g})\end{aligned}$$

where $\mathcal{L}_{\mathcal{M}_{i,j}}(h)$ is defined in equation (3.2.19) with $\{i, j\} \in \mathcal{S}$. Then, with these results, we can build Likelihood Ratio type tests based on $\mathcal{L}_{\mathcal{M}_g}(h; \widehat{W}_\nu^{(T)}, \widehat{B}_{\ell_g}^{(T)})$, or Lagrange Multiplier tests based on $\Delta_{\mathcal{M}_g}(h; \widehat{W}_\nu^{(T)}, \widehat{B}_{\ell_g}^{(T)})$. These tests enjoy the Chernoff-Savage type result described in the next corollary.

Corollary 3.3.1 (Chernoff-Savage Type Result). *Define the σ -field of W_ν as $\mathcal{M}_e := \sigma(W_\nu)$. Then, the Likelihood Ratio type tests and Lagrange Multiplier tests based on \mathcal{M}_g , for any chosen reference densities g_i for $i = 1, \dots, p$, is strictly more efficient than those based on \mathcal{M}_e when f_ν is non-Gaussian and as efficient as the latter when f_ν is Gaussian.*

This result is straightforward, henceforth, its proof is omitted. However, this result is quite meaningful: The usual inference procedures for nonstandard econometric problems are derived under some canonical version of the studied model assuming i.i.d. Gaussian innovations and, as a second step, shown to be valid in a much larger class of models. These canonical widely-used inference procedures are actually, in the limit, the ones based on \mathcal{M}_e , while those constructed using rank statistics are the ones based on \mathcal{M}_g . Thus, Corollary 3.3.1 shows us that by using rank statistics, we can gain efficiency from the non-Gaussian innovation density. To provide a more intuitive expression, the next corollary gives the structural limit experiment

associated to \mathcal{M}_g , denoted by \mathcal{E}_g^* . The associated proof is in Appendix 3.7.

Corollary 3.3.2 (Structural Limit Experiment of \mathcal{M}_g). *The partial sum processes $W_\nu^{(T)}$ and $B_{\ell_g}^{(T)}$ provide us, in the limit, two observed processes $W_\nu \in D^p[0, 1]$ and $B_{\ell_g} \in D^p[0, 1]$. In this limit experiment, they follow the model*

$$dW_\nu(u) = M(u, \theta)du + mdu + dZ_\nu(u), \quad (3.3.8)$$

$$dB_{\ell_g}(u) = J_{f_\nu g} [M(u, \theta) - M(1, \theta)] du + d[Z_{\ell_g}(u) - uZ_{\ell_g}(1)], \quad (3.3.9)$$

where Z_ν and Z_{ℓ_g} are multivariate Brownian motions with covariance matrix in (4.3.15), and where $B^{Z_{\ell_g}}$ is the Brownian bridge of Z_{ℓ_g} .

Actually, if we apply Girsanov's theorem to the model equations of \mathcal{E}_g^* , this leads to the likelihood ratio (or Radon-Nikodym derivative) $\mathcal{L}_{\mathcal{M}_g}(h; W_\nu, B_{\ell_g})$ in (3.2.19). However, $\mathcal{L}_{\mathcal{M}_g}(h; W_\nu, B_{\ell_g})$ contains the so-called “cross Fisher information matrix” $J_{f_\nu g}$, which is unknown and can only be estimated based on some nonparametric estimates of the score functions. The need of nonparametric estimation makes this method hard to implement, and computationally demanding in Monte Carlo simulations. Therefore, we will not apply this method to the two applications in the present paper and we introduce an easier-to-implement method.

Method II: inference procedures based on replacement of B_{ℓ_f} by B_{ℓ_g}

Denote the Brownian bridge of $W_{\ell_{f_\nu}}$ by $B_{\ell_{f_\nu}}$. The “oracle” inference procedures, which attains the semiparametric bounds in Section 3.2.4, can be expressed with W_ν , $B_{\ell_{f_\nu}}$ and J_{f_ν} via the relations in (3.3.1)-(3.3.3). Then, our easier-to-implement method can be described by replacing $B_{\ell_{f_\nu}}$ with B_{ℓ_g} and, subsequently, plugging in $\widehat{W}_\nu^{(T)}$ and $\widehat{B}_{\ell_g}^{(T)}$ as the feasible finite-sample counterparts of W_ν and B_{ℓ_g} . For the standardized Fisher information J_{f_ν} ,

we replace it with an reference Fisher information

$$J_p = \text{diag} (J_{g_1}, \dots, J_{g_p}) , \quad (3.3.10)$$

where J_{g_i} is the scalar Fisher information for marginal reference density g_i (chosen by researchers). This method is much less computationally expensive than Method I. Moreover, it also enjoys some attractive properties.

In Monte Carlo simulations, we find that the Chernoff-Savage type result is also preserved by this method, if all the marginal reference densities g_i are fixed to be Gaussian. We will show this property in the following cointegration and weak instrument applications: the easier-to-implement rank-based statistics by Method II are more powerful than the original residual-based statistics for some non-Gaussian density f , and work as well as them when f is Gaussian.

3.3.4 Optimality with Independent ν_t

In the multivariate case, the (semiparametrically and asymptotically) optimality property in the sense of reaching the semiparametric efficiency bound of the rank-based inferences introduced above are often absent even with correctly specified marginal reference densities. This is due to the fact that although the components of ν_t are uncorrelated, the components of R_t depend on each and other through the (unknown) copula structure. Consequently, some more assumptions need to be imposed in order to claim optimality.

As mentioned in the beginning of this section, one way is to assume that the joint density is elliptical and, as a result, the situation degenerates to the univariate case. However, this assumption is not always realistic for economic and financial data, and it is testable with the estimates of the errors under the null hypothesis. Any deviation from this assumption will break the validity property.

Another possible assumption to impose is that the uncorrelated components of ν_t are actually independent of each other, which, consequently, implies that the score function with respect to a certain direction i equals to

the score function of the marginal density of this direction. Moreover, this assumption implies also that the componentwise ranks are (cross-sectionally) independent. Thus, with correctly specified marginal reference densities $g_i = f_{\nu,i}$, we have $\widehat{B}_{\ell_g}^{(T)} \Rightarrow B_{\ell_{f_\nu}}$, which induces the optimality result.⁸ Nevertheless, even (probably) without the optimality property, we will show in the next two applications that the componentwise-rank-based testing procedure can gain considerable power from non-Gaussian distributions.

3.4 Cointegration

In this section, our new approach is applied to the problem of testing the cointegration rank in a vector autoregressive (VAR) model. The model is set up in Section 3.4.1. Section 3.4.2 derives the semiparametric power envelope of asymptotically invariant tests in a simplified model. A rank-based test is introduced in Section 3.4.3, of which a Monte Carlo study is presented in Section 3.4.4. Section 3.4.5 provides some extensions that are necessary for empirical applications.

3.4.1 The Model

Consider p -dimensional observations $Y_T = (y_1, \dots, y_T)'$ generated by the p -th order vector autoregressive model written in error correction form (ECM)

$$y_t = \mu' d_t + v_t, \quad (3.4.1)$$

$$\Delta v_t = \Pi v_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta v_{t-j} + \varepsilon_t, \quad (3.4.2)$$

for $t = 1, \dots, T$, where Δ denotes first-order differencing, v_{1-k}, \dots, v_0 are deterministic starting values, $\mu' d_t$ stands for a deterministic component where $d_t = 1$ or $d_t = (1, t)'$, and ε_t is a p -dimensional i.i.d. sequence of innovations

⁸Here the optimality result holds for both of the rank-based tests using the two methods above since, with correctly specified marginal reference densities, these two tests rank-based would be identical. Feasibility can be achieved by using nonparametric estimates $\hat{f}_{\nu,i}$ of marginal densities $f_{\nu,i}$, for $i = 1, \dots, p$.

with joint density $f \in \mathcal{F}_p$. In this problem, μ and $\Gamma := (\Gamma_1, \dots, \Gamma_{k-1}) \in \mathbb{R}^{p \times (k-1)p}$ are nuisance parameters to be eliminated, while $\Pi \in \mathbb{R}^{p \times p}$ is the parameter of interest. More precisely, we are interested in knowing the rank r of Π .

3.4.2 Semiparametric Power Envelope

Initially, we begin with a simplified special case, where $d_t = 0$ (henceforth $y_t = v_t$) and $\Gamma = 0$. Employ the local reparameterization $\Pi = C/T$ (see Chan and Wei (1988) and Phillips (1988)), where $C \in \mathbb{R}^{p \times p}$ is fixed while $T \rightarrow \infty$. Then the model becomes

$$\Delta y_t = \frac{C}{T} y_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}. \quad (3.4.3)$$

We are interested in testing the hypothesis

$$H_0 : r = 0 \quad \text{against} \quad H_1 : r > 0,$$

and $r = 0$ if and only if $C = 0$. Due to the lack of knowledge of the true density f , we assume ε_t are generated from density $f_\eta^{(T)}(\varepsilon_t)$ defined in (3.2.3). As mentioned in Section 3.4.1, let $P_{C,\eta}^{(T)}$ denote the law of y_1, \dots, y_T under (3.4.1) with innovation density $f_\eta^{(T)}(\varepsilon_t)$, and $\mathbb{P}_{C,\eta}$ denote the collection of probability measures of the limit experiment. Also, denote the log-likelihood ratio $\log(dP_{C,\eta}^{(T)}/dP_{0,0}^{(T)})$ by $L^{(T)}(C, \eta)$ and the limit representative $\log(d\mathbb{P}_{C,\eta}/d\mathbb{P}_{0,0})$ by $\mathcal{L}(C, \eta)$. The following proposition shows that this cointegration model is of the LABF-type likelihood ratios.

Proposition 3.4.1. *Let $f \in \mathcal{F}_p$, $\eta \in c_{00}$, and $C \in \mathbb{R}^{p \times p}$.*

(a) *Under $P_{0,0}^{(T)}$, we have*

$$L^{(T)}(C, \eta) = \Delta^{(T)}(C, \eta) - \frac{1}{2} Q^{(T)}(C, \eta) + o_{p_0}(1) \quad (3.4.4)$$

with

$$\Delta^{(T)}(C, \eta) = \frac{1}{T} \sum_{t=2}^T (C y_{t-1})' \ell_f(\Delta y_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta' b(\Delta y_t),$$

$$Q^{(T)}(C, \eta) = \frac{1}{T^2} \sum_{t=2}^T (Cy_{t-1})' J_f (Cy_{t-1}) + \frac{2}{T^{3/2}} \sum_{t=1}^T \eta' J_{bf} (Cy_{t-1}) + \eta' \eta.$$

(b) Under $P_{0,0}^{(T)}$, as $T \rightarrow \infty$, we have $L^{(T)}(C, \eta) \Rightarrow \mathcal{L}(C, \eta)$, where

$$\mathcal{L}(C, \eta) = \Delta(C, \eta) - \frac{1}{2} \mathcal{Q}(C, \eta) \quad (3.4.5)$$

with

$$\begin{aligned} \Delta(C, \eta) &= \int_0^1 [CW_\varepsilon(u)]' dW_{\ell_f}(u) + \eta' W_b(1), \\ \mathcal{Q}(C, \eta) &= \int_0^1 [CW_\varepsilon(u)]' J_f CW_\varepsilon(u) du + 2\eta' \int_0^1 J_{bf} CW_\varepsilon(u) du + \eta' \eta. \end{aligned}$$

(c) Under $\mathbb{P}_{0,0}$, $\exp(\mathcal{L}(C, \eta))$ has zero expectation.

The proof is provided in Appendix 3.7. By Theorem 3.2.1, we can get immediately the structural limit experiment $\mathcal{E}(f)$ with $M(u, \theta)$ replaced by $CW_\varepsilon(u)$ and $m = 0$ in (3.2.12), (3.2.14) and (3.2.13). From the discussion in Remark 3.2.1, the (sufficient) structural limit experiment $\mathcal{E}^*(f)$ is

$$dW_\varepsilon(u) = CW_\varepsilon(u) du + dZ_\varepsilon(u), \quad (3.4.6)$$

$$dW_{\ell_f}(u) = J_f CW_\varepsilon(u) du + J_{fb} \eta du + dZ_{\ell_f}(u). \quad (3.4.7)$$

By Theorem 3.2.2 and Remark 3.2.1, in the structural limit experiment $\mathcal{E}^*(f)$, the maximal invariant is $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$. The log-likelihood ratio of $\mathcal{M}_{0,\eta}^*$ in the limit experiment is given by

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C) = E[\mathcal{L}(C, \eta) | \mathcal{M}_{0,\eta}^*] = \Delta_{\mathcal{M}_{0,\eta}^*}(C) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(C) \quad (3.4.8)$$

with

$$\begin{aligned} \Delta_{\mathcal{M}_{0,\eta}^*}(C) &= \int_0^1 [CW_\varepsilon(u)]' dB_{\ell_f}(u) + \int_0^1 [CW_\varepsilon(u)]' du \Sigma^{-1} W_\varepsilon(1) \\ \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(C) &= \int_0^1 [CW_\varepsilon(u)]' J_f CW_\varepsilon(u) du \\ &\quad + \int_0^1 [CW_\varepsilon(u)]' du (\Sigma^{-1} - J_f) \int_0^1 CW_\varepsilon(u) du. \end{aligned}$$

Detailed calculations are provided in Appendix 3.8. Define $\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C)$ -based likelihood test $\phi(C, \alpha) := \mathbb{1}\{\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C) > \kappa_\alpha(C)\}$ where $\kappa_\alpha(C)$ is the $1 - \alpha$ quantile of $\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C)$. It follows from the Neyman-Pearson lemma that the upper power bound of tests invariant with respect to η in the limit experiment is

$$\Psi(C, \alpha) = \mathbb{E} \left[\phi(C, \alpha) \frac{d\mathbb{P}_{C,\eta}}{d\mathbb{P}_{0,0}} \right]. \quad (3.4.9)$$

The Asymptotic Representation Theorem yields the following main result on the asymptotic power envelope.

Theorem 3.4.1 (Semiparametric Power Envelope). *Let $f \in \mathcal{F}_p$ and $\alpha \in (0, 1)$. Let $\phi_T(y_1, \dots, y_T)$ be an asymptotically invariant test of size α , i.e., $\limsup_{T \rightarrow \infty} \mathbb{E}_{0,\eta} \phi_T \leq \alpha$ for all $\eta \in c_{00}$. Denote the power of test ϕ_T , $\mathbb{E}_{C,\eta}[\phi_T]$, by $\pi_T(C, \eta)$. Then we have*

$$\limsup_{T \rightarrow \infty} \pi_T(C, \eta) \leq \Psi(C, \eta)$$

for any $C \in \mathbb{R}^{p \times p}$ and $\eta \in c_{00}$.

The concept of semiparametric power envelope is used a bit early here, since it is nothing but an upper bound without showing it is attainable. For this purpose, similarly as Jansson (2008) for the unit root testing problem, we use a \hat{f} -based inference to claim the attainability.⁹ The \hat{f} -based statistic, in this case, is by plugging in where appear the score function ℓ_f and Fisher information J_f with their estimates based on some consistent estimate \hat{f} . Rewrite the first term of $\Delta_{\mathcal{M}_{0,\eta}^*}(C)$ as

$$\int_0^1 [CW_\varepsilon(u)]' dB_{\ell_f}(u) = \int_0^1 [C\widetilde{W}_\varepsilon(u)]' dW_{\ell_f}(u),$$

⁹Unlike the unit root case in Zhou, van den Akker, and Werker (2016), rank-based inference does not attain the semiparametric efficiency bound with $g = f$ (or $g = \hat{f}$) in multivariate problems for the reasons discussed in Section 3.3.4 (or, unless make the independence assumption therein).

where $\widetilde{W}_\varepsilon(u) = W_\varepsilon(u) - \int_0^1 W_\varepsilon(u)du$. The \hat{f} -based version of this part is $T^{-1} \sum_{t=2}^T (C\widetilde{y}_{t-1})' \hat{\ell}_f(\Delta y_t)$, where $\widetilde{y}_{t-1} = y_{t-1} - \frac{1}{T} \sum_{t=2}^T y_{t-1}$. As a consequence, even if there is no \sqrt{T} -unbiased estimator for ℓ_f when f is unrestricted (see Klaassen (1987)), the bias existing in estimating ℓ_f will be canceled out since $T^{-1} \sum_{t=2}^T (C\widetilde{y}_{t-1})'[a + \ell_f(\Delta y_t)] = T^{-1} \sum_{t=2}^T (C\widetilde{y}_{t-1})'\ell_f(\Delta y_t)$ for any $a \in \mathbb{R}^p$. Thus, the “plug-in” version of $\phi_T(C, \eta)$ with consistent estimates of ℓ_f and J_f attains the upper bound $\Psi(C, \eta)$ without breaking its validity, which make $\Psi(C, \eta)$ indeed the semiparametric power envelope.

3.4.3 Rank-based Tests

In this section, we propose a rank-based version of the Johansen trace test (c.f., Johansen (1991)). First, based on the form of maximal invariant's likelihood ratio in (3.4.8) and Method II proposed in Section 3.3. First we rewrite $\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C)$ using W_ν and $B_{\ell_{f_\nu}}$ as

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(\tilde{C}) = \Delta_{\mathcal{M}_{0,\eta}^*}(\tilde{C}) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(\tilde{C}) \quad (3.4.10)$$

where

$$\begin{aligned} \Delta_{\mathcal{M}_{0,\eta}^*}(\tilde{C}) &= \int_0^1 [\tilde{C}W_\nu(u)]' dB_{\ell_{f_\nu}}(u) + \int_0^1 [\tilde{C}W_\nu(u)]' du W_\nu(1) \\ \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(\tilde{C}) &= \int_0^1 [\tilde{C}W_\varepsilon(u)]' J_{f_\nu} \tilde{C}W_\varepsilon(u) du \\ &\quad + \int_0^1 [\tilde{C}W_\nu(u)]' du (I_p - J_{f_\nu}) \int_0^1 \tilde{C}W_\nu(u) du \end{aligned}$$

with $\tilde{C} = \Sigma^{-\frac{1}{2}'} C \Sigma^{\frac{1}{2}}$. Then, following Method II, we replace W_ν , $B_{\ell_{f_\nu}}$ and J_{f_ν} by $\widehat{W}_\nu^{(T)}$, $\widehat{B}_{\ell_g}^{(T)}$ and J_p , respectively. Then we get the componentwise-rank based likelihood-ratio statistic

$$\widehat{L}_g^{(T)}(\tilde{C}) = \widehat{\Delta}_g^{(T)}(\tilde{C}) - \frac{1}{2} \widehat{Q}_g^{(T)}(\tilde{C}), \quad (3.4.11)$$

where

$$\begin{aligned} \widehat{\Delta}_g^{(T)}(\tilde{C}) &= \int_0^1 [\tilde{C}\widehat{W}_\nu^{(T)}(u-)]' d\widehat{B}_{\ell_g}^{(T)}(u) + \int_0^1 [\tilde{C}\widehat{W}_\nu^{(T)}(u-)]' du \widehat{W}_\nu^{(T)}(1), \\ \widehat{Q}_g^{(T)}(\tilde{C}) &= \int_0^1 [\tilde{C}\widehat{W}_\nu^{(T)}(u-)]' J_p \tilde{C}\widehat{W}_\nu^{(T)}(u-) du \end{aligned}$$

$$+ \int_0^1 [\tilde{C}\widehat{W}_\nu^{(T)}(u-)]' du (I_p - J_p) \int_0^1 \tilde{C}\widehat{W}_\nu^{(T)}(u-) du.$$

In the spirit of Johansen (1991), the Rank-Johansen test statistic is thus

$$\widehat{LR}_g^{(T)} = \text{tr} \left[\widehat{\Delta}_g' \widehat{Q}_g^{-1} \widehat{\Delta}_g \right], \quad (3.4.12)$$

where

$$\begin{aligned} \widehat{\Delta}_g &= \int_0^1 \widehat{W}_\nu^{(T)}(u-) d\widehat{B}_{\ell_g}^{(T)}(u)' + \int_0^1 \widehat{W}_\nu^{(T)}(u-) du \widehat{W}_\nu^{(T)}(1)', \\ \widehat{Q}_g &= J_p \int_0^1 \widehat{W}_\nu^{(T)}(u-) \widehat{W}_\nu^{(T)}(u-)' du + (I_p - J_p) \int_0^1 \widehat{W}_\nu^{(T)}(u-) du \int_0^1 \widehat{W}_\nu^{(T)}(u-)' du. \end{aligned}$$

Denote by \mathcal{LR}_g the limit of the above rank-based Johansen trace statistic $\widehat{LR}_g^{(T)}$ and it is straightforwardly given by

$$\mathcal{LR}_g = \text{tr} \left[\Delta_g' \mathcal{Q}_g^{-1} \Delta_g \right]$$

with

$$\begin{aligned} \Delta_g &= \int_0^1 W_\nu(u) dB_{\ell_g}(u)' + \int_0^1 W_\nu(u) du W_\nu(1)', \\ \mathcal{Q}_g &= J_p \int_0^1 W_\nu(u) W_\nu(u)' du + (I_p - J_p) \int_0^1 W_\nu(u) du \int_0^1 W_\nu(u)' du. \end{aligned}$$

The critical value function $\kappa_g(\alpha)$ is taken as $(1 - \alpha)$ quantile of \mathcal{LR}_g with significance level α . This leads to the feasible test

$$\phi_g^{(T)} = \mathbb{1} \left\{ \widehat{LR}_g^{(T)} \geq \kappa_g(\alpha) \right\}, \quad (3.4.13)$$

where “ $\mathbb{1}$ ” is the indicator function.

3.4.4 Monte Carlo Simulations

Monte Carlo setup

In this Monte Carlo study, we consider the two-dimension model ($p = 2$) of the simplified form in (3.4.3), with i.i.d. innovations ε_t generated from some density $f \in \mathcal{F}_2$. The local parameter of interest C is specified by the following form

$$C = c \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

where $c \in (-\infty, 0]$. So, $c = 0$ implies the null hypothesis $r = 0$. Here we fix the angle of the matrix C , but try different values for the correlation $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ of the two innovation series. We simulate 20,000 independent replications from this data generating process with sample sizes $T = 2500$ for the large sample case and $T = 250$ for the small sample case. In order to reduce computational burden, the critical values for our rank-based Johansen test is simulated and fixed beforehand for each case so that we do not need to simulate it for each iteration. The significant level is chosen to be 5% for all cases.

Large-sample results

In Figure 3.1, we try four combinations of different true joint density and marginal reference densities with fixed $\rho = 0$. Specifically, when f is multivariate t_3 while g_1 and g_2 are univariate t_3 , the power function of our rank-based Johansen test is much higher than that of the Johansen test. Moreover, it is very close to the semiparametric power envelope, which indicates that the efficiency gain mainly comes from the marginal densities rather than the copula structure.¹⁰

The traditional Chernoff-Savage property of rank-based inference also holds here: keep g_1 and g_2 to be Gaussian, our rank-based Johansen test can gain considerable power when f is multi- t_3 comparing to the Johansen test, and works as well as the Johansen test when f is Gaussian. Unreported simulation results show similar results hold for all other (tried) non-Gaussian f .

However, when f is Gaussian (or close to Gaussian), if we choose some marginal reference densities that are “far from” Gaussian, our test will be of less power than the Johansen test, see the last sub-figure where f is Gaussian and g_1 and g_2 are t_3 . Thus, as a guidance of choosing g_i for $i = 1, \dots, p$, an

¹⁰Maybe all the efficiency gain comes from the marginal densities since here g_i is not correctly specified (only nearly correctly specified), since the marginal density of multivariate t_3 is not univariate t_3 .

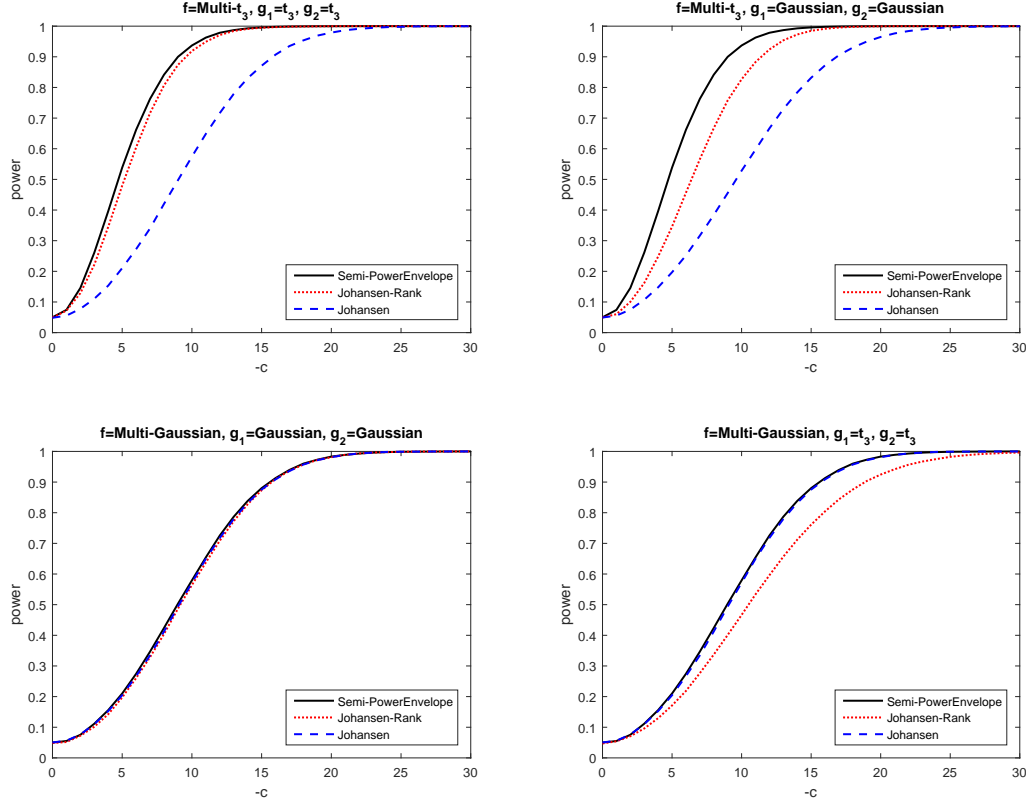


Figure 3.1: Power functions of the Johansen test and our Rank-based Johansen test.

conservative strategy would be choosing them to be Gaussian all the time.

In Figure 3.2, we keep f to be Multi- t_3 , g_1 and g_2 to be Gaussian, and try different values of ρ to show the componentwise-rank-procedure proposed in Section 3.3 works well with sectionally-correlated innovations. In all cases with $\rho = \pm 0.25, \pm 0.75$, the rank-based Johansen test dominates the Johansen test with significantly more power.

Small-sample results

Figure 3.3 is actually the small-sample version of the Figure 3.2 with sample size $T = 250$ and everything else being the same. As we can see in the figure,

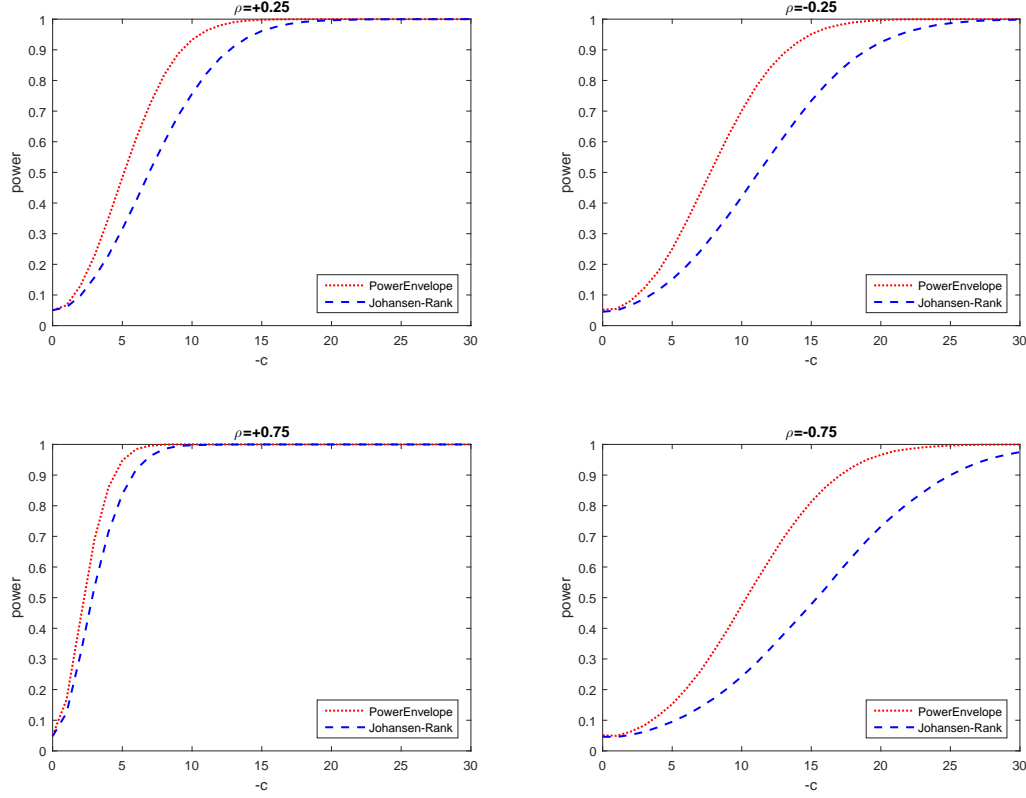


Figure 3.2: Power functions of the Johansen test and the Rank-based Johansen test.

our rank-based Johansen test works well in the sense that the sizes are close to 5%, and the power functions are much higher than those of the Johansen test.

3.4.5 Extensions

In this section, we address some extensions from the current simplified model, which is useful for empirical applications. First, we include the deterministic component part d_t . Then, we consider the problem of testing more general hypotheses on matrix Π , to be specific, testing the null hypothesis that Π is of rank $r = r_0$ (for some $r_0 < p$) against the alternative

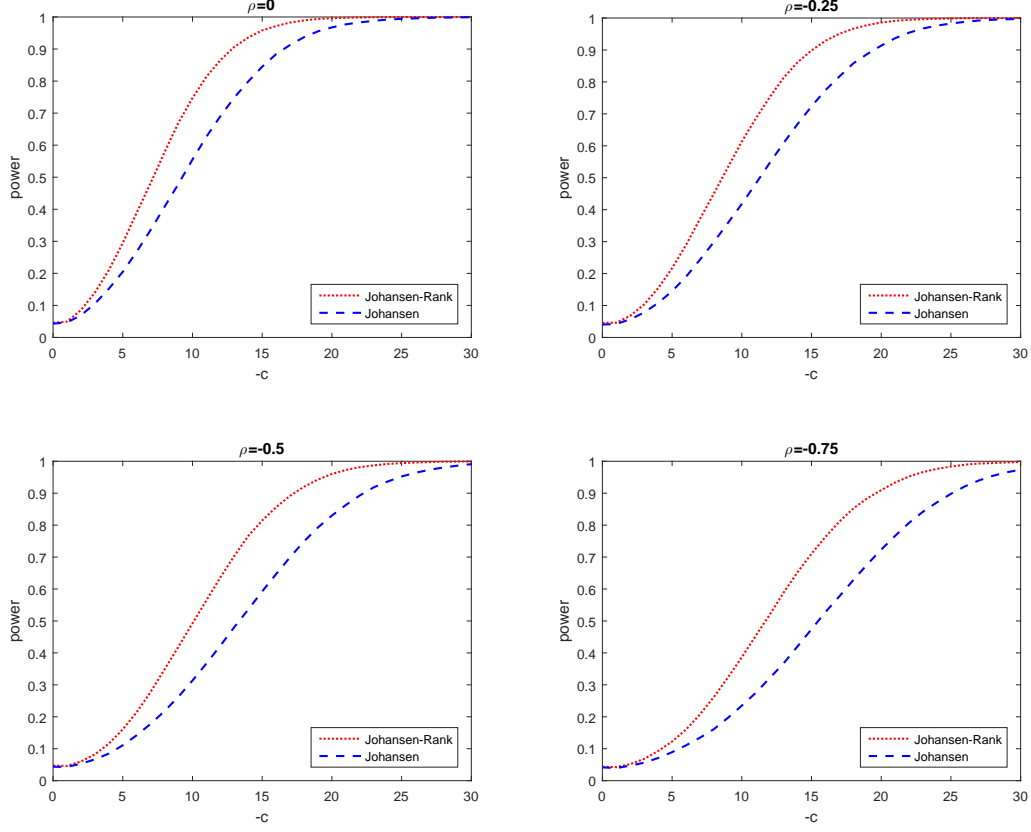


Figure 3.3: Power functions of the Johansen test and the Rank-based Johansen test.

$r > r_0$.

Deterministic component

We divide into two cases here as in Section 3.4.1: $d_t = 1$ with $\mu = \mu_1 \in \mathbb{R}^p$, and $d_t = (1, t)'$ with $\mu = (\mu_1, \mu_2) \in \mathbb{R}^{p \times 2}$, since they would have different limits.

The model with $d_t = 1$ shares the same limit as the simplified model, since this extension only adds an constant μ_1 to the observations y_t of the simplified case above, and this constant μ_1 will be omitted as $T \rightarrow \infty$. On

the other hand, since our current rank-based testing statistic depends only through the increments of the observations Δy_t and its associated vector of ranks, which are all invariant with respect to the transformation of adding a constant to y_t , therefore, this type of deterministic part will not lead to any difference on the sequence side as well.

For the model with $d_t = (1, t)'$, the limit changes because the coefficient for t in the deterministic term, μ_2 , will not disappear as $T \rightarrow \infty$. In this case, the log likelihood ratio admits an expansion of the form

$$L^{(T)}(C, \mu_2, \eta) = \Delta^{(T)}(C, \mu_2, \eta) - \frac{1}{2}Q^{(T)}(C, \mu_2, \eta) + o_{p_0}(1)$$

where

$$\begin{aligned} \Delta^{(T)}(C, \mu_2, \eta) &= \frac{1}{T} \sum_{t=2}^T (Cy_{t-1})' \ell_f(\Delta y_t) + \frac{1}{T} \sum_{t=2}^T (d_{C,t} \mu_2)' \ell_f(\Delta y_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta' b(\Delta y_t), \\ Q^{(T)}(C, \mu_2, \eta) &= \frac{1}{T^2} \sum_{t=2}^T (Cy_{t-1})' J_f(Cy_{t-1}) + \frac{1}{T^2} \sum_{t=2}^T (d_{C,t} \mu_2)' J_f(d_{C,t} \mu_2) \\ &\quad + \frac{2}{T^2} \sum_{t=2}^T (Cy_{t-1})' J_f(d_{C,t} \mu_2) + \frac{2}{T^{3/2}} \sum_{t=1}^T \eta' J_{bf}(Cy_{t-1}) \\ &\quad + \frac{2}{T^{3/2}} \sum_{t=1}^T \eta' J_{bf}(d_{C,t} \mu_2) + \eta' \eta. \end{aligned}$$

with $d_{C,t} = I_p - C(t-1)/T$. It weakly converges to

$$\mathcal{L}(C, \mu_2, \eta) = \Delta(C, \mu_2, \eta) - \frac{1}{2}\mathcal{Q}(C, \mu_2, \eta)$$

with

$$\begin{aligned} \Delta(C, \mu_2, \eta) &= \int_0^1 [CW_\varepsilon(u)]' dW_{\ell_f}(u) + \int_0^1 (d_{C,u} \mu_2)' dW_{\ell_f}(u) + \eta' W_b(1), \\ \mathcal{Q}(C, \mu_2, \eta) &= \int_0^1 [CW_\varepsilon(u)]' J_f CW_\varepsilon(u) du + \int_0^1 (d_{C,u} \mu_2)' J_f (d_{C,u} \mu_2) du \\ &\quad + 2 \int_0^1 [CW_\varepsilon(u)]' J_f (d_{C,u} \mu_2) du + 2\eta' \int_0^1 J_{bf} CW_\varepsilon(u) du \\ &\quad + 2\eta' \int_0^1 J_{bf} d_{C,u} \mu_2 du + \eta' \eta, \end{aligned}$$

where $d_{C,u} := I_p - Cu$. Then, the sufficient structural limit experiment $\mathcal{E}_d^*(f)$ is given by

$$\begin{aligned} dW_\varepsilon(u) &= CW_\varepsilon(u)du + (I_p - Cu)\mu_2 du + dZ_\varepsilon(u), \\ dW_{\ell_f}(u) &= J_f CW_\varepsilon(u)du + J_f(I_p - Cu)\mu_2 du + J_{fb}\eta du + dZ_{\ell_f}(u). \end{aligned}$$

The maximal invariant is given by $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$, and its log likelihood ratio is

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2) = \Delta_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2)$$

with

$$\begin{aligned} \Delta_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2) &= \int_0^1 [CW_\varepsilon(u) + (I_p - Cu)\mu_2]' d[B_{\ell_f}(u) + \Sigma^{-1}W_\varepsilon(1)] \\ \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2) &= \int_0^1 [CW_\varepsilon(u) + (I_p - Cu)\mu_2]' J_f [CW_\varepsilon(u) + (I_p - Cu)\mu_2] du \\ &\quad + \int_0^1 [CW_\varepsilon(u) + (I_p - Cu)\mu_2]' du (\Sigma^{-1} - J_f) \int_0^1 [CW_\varepsilon(u) + (I_p - Cu)\mu_2] du. \end{aligned}$$

Then, μ_2 can be eliminated by profiling the likelihood function as

$$\mathcal{L}_*^d(C) = \max_{\mu_2} \mathcal{L}_{\mathcal{M}_{0,\eta}^*}^d(C, \mu_2) - \mathcal{L}_{\mathcal{M}_{0,\eta}^*}^d(C, 0).$$

An explicit form of $\mathcal{L}_*^d(C)$ is given by

$$\mathcal{L}_*^d(C) = \Delta_c(C) - \frac{1}{2} \mathcal{Q}_c(C) + \frac{1}{2} \mathcal{L}_d(C) - \frac{1}{2} \mathcal{L}_d(0)$$

with $\mathcal{L}(C) = \Delta_d(C) \mathcal{Q}_{dd}^{-1}(C) \Delta_d(C) - 2 \mathcal{Q}_{cd}(C) \mathcal{Q}_{dd}^{-1}(C) \Delta_d(C) + \mathcal{Q}_{cd}(C) \mathcal{Q}_{dd}^{-1}(C) \mathcal{Q}_{cd}(C)$,

where

$$\begin{aligned} \Delta_c(C) &= \int_0^1 [CW_\varepsilon(u)]' d[B_{\ell_f}(u) + \Sigma^{-1}W_\varepsilon(1)], \\ \Delta_d(C) &= \int_0^1 (I_p - Cu)' d[B_{\ell_f}(u) + \Sigma^{-1}W_\varepsilon(1)], \\ \mathcal{Q}_{cd}(C) &= \int_0^1 [CW_\varepsilon(u)]' J_f (I_p - Cu) du + \int_0^1 [CW_\varepsilon(u)]' du (\Sigma^{-1} - J_f) (I_p - C), \\ \mathcal{Q}_{dd}(C) &= \int_0^1 (I_p - Cu)' J_f (I_p - Cu) du + (I_p - C)' (\Sigma^{-1} - J_f) (I_p - C). \end{aligned}$$

Similarly as in Section 3.4.2, the semiparametric power envelope for the alternative point C is defined by the power function of the likelihood ratio

test based on $\mathcal{L}_*^d(C)$. The rank-based statistic can be constructed by mimicking the form of Johansen test as in (3.4.12) based on the likelihood $\mathcal{L}_*^d(C)$ with detrended versions of Δy_t and their ranks.

Reduced rank hypothesis

Suppose there already are r_0 combinations of p nonstationary that are stationary, and we want to know if there is another one more combination. Then the hypotheses of interested are

$$H_0 : r = r_0, \text{ against } H_1 : r > r_0,$$

Now consider the parameterization of Π for this case

$$\Pi = \alpha\beta' + \frac{\alpha_\perp C \alpha_\perp'}{T}$$

where $C \in \mathbb{R}^{p-r_0 \times p-r_0}$ is the parameter of interest, while $\alpha \in \mathbb{R}^{p \times r_0}$, $\alpha_\perp \in \mathbb{R}^{p \times p-r_0}$ and $\beta \in \mathbb{R}^{p \times r_0}$ are nuisance parameters. Moreover, α and α_\perp are orthogonal.

Since $r = r_0$ if and only if $C = 0$, the problem becomes testing $C = 0$ with transformed data $\alpha_\perp' y_t$. Consistent estimation of α_\perp is satisfied when α and β can be consistently estimated (see Section 2.3 of Boswijk, Jansson, and Nielsen (2015)). Moreover, according to the supplement materials of Hallin, Van den Akker, and Werker (2016), the testing inference for C is adaptive with respect to the estimations of α and β in the sense that the score of C and the scores of local parameters of α and β are uncorrelated. Thus, the procedure can be proceed as: (i) estimate α and β using the original data y_t , (ii) obtain the estimator $\hat{\alpha}_\perp$ for α_\perp using these results and orthogonality of α and α_\perp ,¹¹ and (iii) apply the rank-based testing procedure on the transformed data $\hat{\alpha}_\perp y_t$.

¹¹For the details of this step, see also Section 2.3 of Boswijk, Jansson, and Nielsen (2015).

3.5 Weak Instrument

In this section, our new approach is applied to the problem of instrumental variable regression with several weak instruments. We set up the model in Section 3.5.1. The semiparametric power envelope is derived in Section 3.5.2. In Section 3.5.3, we provide limit representations of the currently-used tests in the Gaussian case, and extend them to the general density function case. We provide the associated rank-based versions of these tests in Section 3.5.4 and their simulation results in Section 3.5.5.

3.5.1 Model

As in Andrews, Moreira, and Stock (2006) (subsequently abbreviated AMS), we consider the problem of making statistical inference for the coefficient of a scalar endogenous variable with the presence of weak instruments in the structural model

$$\begin{aligned} y_{1t} &= y_{2t}\beta + x_t'\xi + v_t, \\ y_{2t} &= z_t'\pi + x_t'\gamma_2 + \varepsilon_{2t}, \end{aligned} \tag{3.5.1}$$

for $t = 1, \dots, T$, where y_{1t} and y_{2t} are two scalar endogenous variables; $x_t \in \mathbb{R}^q$ and $z_t \in \mathbb{R}^k$ are exogenous variables which are observable, non-stochastic, and satisfy Assumption 3.2.2; and u, v_2 are unobserved error terms. In this model, the exogenous variables z_t are employed as instruments, whose correlations with the endogenous variable y_{2t} are denoted by π . The reduced form equations are given by

$$\begin{aligned} y_{1t} &= z_t'\pi\beta + x_t'\gamma_1 + \varepsilon_{1t}, \\ y_{2t} &= z_t'\pi + x_t'\gamma_2 + \varepsilon_{2t}, \end{aligned} \tag{3.5.2}$$

where $\gamma_1 = \xi + \gamma_2\beta$ and $\varepsilon_{1t} = v_t + \varepsilon_{2t}\beta$. Assume $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a two-dimensional i.i.d. sequence of innovations with joint density $f \in \mathcal{F}_2$. We are interested in testing

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0. \tag{3.5.3}$$

3.5.2 Semiparametric Bound with Given π

In this linear regression model, $x_{1t} = 1$ and the associated coefficient $\gamma_{1,1}$ is regarded as the nuisance constant term, hence, we set μ equal to zero (accordingly, m is known to be zero) for the reason that it is not identifiable from $\gamma_{1,1}$. Moreover, in Assumption 3.2.2, we assume that z_t and x_t are (asymptotically) independent.¹² This leads to the fact that $\mu_z = 0$, which induces the adaptive result (for β) later. For asymptotic analysis, we impose the following weak IV fixed alternative (WIV-FA) asymptotics of Staiger and Stock (1997).

Assumption 3.5.1 (WIV-FA). For some non-stochastic k -vector c , $\pi = c/\sqrt{T}$. β is fixed for all $T \geq 1$.

Under Assumption 3.5.1, Cattaneo, Crump, and Jansson (2012) provides an adaptation result for one-step estimators of γ_1 and γ_2 . This means that we can eliminate γ_1 and γ_2 by plugging in consistent estimators without affecting the (asymptotic) results for the remaining parameters. Therefore, we delete the terms $x'_t\gamma_1$ and $x'_t\gamma_2$ in the equation system (3.5.2) without loss of generality, since we could always transform the data by $\tilde{y}_{1t} = y_{1t} - x'_{1t}\hat{\gamma}_1$ and $\tilde{y}_{2t} = y_{2t} - x'_{2t}\hat{\gamma}_2$ where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are some consistent estimators, and use the newly generated data \tilde{y}_{1t} and \tilde{y}_{2t} . Thus, we can further simplify the model to

$$\begin{aligned} y_{1t} &= z'_t\pi\beta + \varepsilon_{1t}, \\ y_{2t} &= z'_t\pi + \varepsilon_{2t}. \end{aligned} \tag{3.5.4}$$

The following proposition gives the LAN result for this model. The proof is given in Appendix 3.7.

¹²This causes no loss of generality since we can always replace z_t by the residuals of a linear regression of the original z_t on x_t .

Proposition 3.5.1. *Let $f \in \mathcal{F}_2$, $\eta \in c_{00}$, and let Assumption 3.5.1 hold.*

Define $a(\beta) = \begin{pmatrix} \beta \\ 1 \end{pmatrix}$, then we have

(a) *Under $P_{0,0,0}^{(T)}$, the log-likelihood ratio function can be expanded as*

$$L^{(T)}(\beta, c, \eta) = \Delta^{(T)}(\beta, c, \eta) - \frac{1}{2} \mathcal{Q}(\beta, c, \eta) + o_{p_0}(1) \quad (3.5.5)$$

with

$$\Delta^{(T)}(\beta, c, \eta) = \frac{1}{\sqrt{T}} \sum_{t=1}^T a(\beta)'(c' z_t) \ell_f(y_{1t}, y_{2t}) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta' b(y_{1t}, y_{2t}),$$

$$\mathcal{Q}(\beta, c, \eta) = a(\beta)'(c' D_{zz} c) J_f a(\beta) + 2a(\beta)' J_{fb} \eta c' \mu_z + \eta' \eta.$$

(b) *As $T \rightarrow \infty$, the efficient score $\Delta^{(T)}(\beta, c, \eta)$ weakly converges to*

$$\Delta(\beta, c, \eta) = \int_0^1 a(\beta)'(c' z_u) dW_{\ell_f}(u) + \eta' W_b(1),$$

which, under the null hypothesis, is a normally distributed random variable with mean zero and variance $\mathcal{Q}(\beta, c, \eta)$.

(c) *Define the limit log-likelihood ratio by $\mathcal{L}(\beta, c, \eta) = \Delta(\beta, c, \eta) - \frac{1}{2} \mathcal{Q}(\beta, c, \eta)$.*

Under $P_{0,0,0}$, $\exp(\mathcal{L}(\beta, c, \eta))$ has zero expectation.

This proposition indicates that, if we regard parameter c as known, the limit experiment is of the conventional LAN structure, in which the observation is one draw of a normal random variable. This type of limit experiment is well addressed by the literature and normally there is little point pursuing a structural version of it. Nevertheless, in reality, c is not known, and the structural limit experiment may show us the way to eliminate it. This is exactly what is next: the structural limit experiment tells us how to eliminate nuisance parameter c , the results in the Gaussian case coincide with the existing literature, and we extend the results to the general innovation density case.

Likewise, Girsanov's theorem gives the structural limit experiment $\mathcal{E}^*(f)$, in which the observation processes W_ε and W_{ℓ_f} is generated by the model

$$dW_\varepsilon(u) = a(\beta)(c'z_u)du + dZ_\varepsilon(u), \quad (3.5.6)$$

$$dW_{\ell_f}(u) = J_f a(\beta)(c'z_u)du + J_{fb}\eta du + dZ_{\ell_f}(u), \quad (3.5.7)$$

where Z_ε and Z_{ℓ_f} are zero-drift Brownian motions with covariance given in (4.2.15). Since now we have $m = 0$, Theorem 3.2.2 shows that $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$ is the maximal invariant w.r.t the transformation $\mathfrak{g}_{0,\eta}$. The log-likelihood ratio of $\mathcal{M}_{0,\eta}^*$ is

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(\beta, c) = \Delta_{\mathcal{M}_{0,\eta}^*}(\beta, c) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(\beta, c), \quad (3.5.8)$$

where

$$\begin{aligned} \Delta_{\mathcal{M}_{0,\eta}^*}(\beta, c) &= a(\beta)' \int_0^1 c' z_u dB_{\ell_f}(u), \\ \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(\beta, c) &= a(\beta)' [(c' D_{zz} c) - (c' \mu_z)^2] J_f a(\beta). \end{aligned}$$

The semiparametric bound can be induced by the efficient score function $\Delta_{\mathcal{M}_{0,\eta}^*}$. The fact that $\mu_z = 0$ implies $a(\beta)' \int_0^1 c' z_u dB_{\ell_f}(u) = a(\beta)' \int_0^1 c' z_u dW_{\ell_f}(u)$, and then $\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(\beta, c) = \mathcal{L}(\beta, c, 0)$. Therefore, the semiparametric power bound equals to the parametric power bound where η is known to be zero, which is no other than the adaptivity result of β when we treat c as known. This adaptivity result is found in Cattaneo, Crump, and Jansson (2012), which follows immediately from Bickel (1982)'s result on adaptive estimation of slope coefficients in a regression model.

3.5.3 Semiparametric Efficient Tests with Known f

Gaussian case

Consider the case that ε_t is normally distributed, where the score function ℓ_f is $\Sigma^{-1}\varepsilon_t$. This implies that $W_{\ell_f} = \Sigma^{-1}W_\varepsilon$ and, as a result, no process other than W_ε is observed in the limit experiment. We express the structural limit experiment equation (3.5.6) element-wisely as

$$dW_{\varepsilon_1}(u) = \beta c' z_u du + dZ_{\varepsilon_1}(u),$$

$$dW_{\varepsilon_2}(u) = c' z_u du + dZ_{\varepsilon_2}(u).$$

In this limit experiment, our goal is to test the null hypothesis $\beta = \beta_0$ against the alternative $\beta \neq \beta_0$, where c is treated as a nuisance parameter. However, it is not straightforward to eliminate c by employing some restrictions. Nevertheless, we can eliminate it under the null if our test statistics is based on $dW_{\varepsilon_1} - \beta_0 dW_{\varepsilon_2}$. Following this idea, we define the statistic

$$\mathcal{S}_\phi = D_{zz}^{-1/2} \int_0^1 z_u d[W_\varepsilon(u)' b_0] \cdot (b_0' \Sigma b_0)^{-1/2}, \quad b_0 = \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix}. \quad (3.5.9)$$

Since here z_u is a deterministic process, \mathcal{S}_ϕ is a k -dimensional normal random variable with mean $D_{zz}^{1/2} c(\beta - \beta_0) (b_0' \Sigma b_0)^{-1/2}$ and identity covariance matrix I_k . A reasonable test statistics would be

$$\mathcal{AR}_\phi = \mathcal{S}_\phi' \mathcal{S}_\phi, \quad (3.5.10)$$

which is $\chi^2(k)$ distributed under H_0 , and we reject the null when it is large enough. The finite-sample representative of \mathcal{AR}_ϕ is the AR test statistic introduced by Anderson, and Rubin (1949), $AR_\phi = S_\phi' S_\phi$ with $S_\phi = (Z_T' Z_T)^{-1/2} Z_T' Y_T b_0 (b_0' \Sigma b_0)^{-1/2}$.

However, due to lack of knowledge about c , the AR test suffers loss of power when $k > 1$. Intuitively, the power coming from the “drift difference” between W_{ε_1} and $\beta_0 W_{\varepsilon_2}$ spreads into k dimensions. To avoid this loss of power, we recall the score function of c under the null hypothesis

$$\int_0^1 z_u d[a_0' \Sigma^{-1} W_\varepsilon(u)], \quad \text{with } a_0 = a(\beta_0) = \begin{pmatrix} \beta_0 \\ 1 \end{pmatrix}.$$

Here W_{ℓ_f} is replaced by $\Sigma^{-1} W_\varepsilon$ because f is Gaussian. Define the following one-to-one function of the score above as

$$\mathcal{T}_\phi = D_{zz}^{-1/2} \int_0^1 z_u d[W_\varepsilon(u)' \Sigma^{-1} a_0] \cdot (a_0' \Sigma^{-1} a_0)^{-1/2}. \quad (3.5.11)$$

\mathcal{T}_ϕ is normally distributed with mean $D_{zz}^{1/2} c[a(\beta) \Sigma^{-1} a_0] (a_0' \Sigma^{-1} a_0)^{-1/2}$ and identity covariance matrix I_k . The corresponding finite-sample representative of \mathcal{T}_ϕ is given by $T_\phi = (Z_T' Z_T)^{-1/2} Z_T' Y_T \Sigma^{-1} a_0 (a_0' \Sigma^{-1} a_0)^{-1/2}$ (c.f.

AMS), which plays the role of a consistent estimator of π with sample size T . Thus, replacing the unknown parameter π by T_ϕ , we have the associated efficient score $S'_\phi T_\phi$ and, sequentially, the Lagrange Multiplier (LM) test statistic $LM_\phi = (S'_\phi T_\phi)^2 / T'_\phi T_\phi$ introduced by Kleibergen (2002). As the asymptotic counterpart, the limit Lagrange Multiplier test statistic is

$$\mathcal{LM}_\phi = \frac{(S'_\phi \mathcal{T}_\phi)^2}{\mathcal{T}'_\phi \mathcal{T}_\phi}. \quad (3.5.12)$$

Similarly, the limit representative of the Conditional Likelihood Ratio test (CLR) introduced by Moreira (2003) is

$$\mathcal{CLR}_\phi = \frac{1}{2} \left[S'_\phi \mathcal{S}_\phi - \mathcal{T}'_\phi \mathcal{T}_\phi + \sqrt{(S'_\phi \mathcal{S}_\phi - \mathcal{T}'_\phi \mathcal{T}_\phi)^2 + 4(S'_\phi \mathcal{T}_\phi)^2} \right]. \quad (3.5.13)$$

General case

In the general case we relax the Gaussian assumption on the density f , assuming only $f \in \mathcal{F}_2$. The associated structural limit experiment based on $\mathcal{M}_{0,\eta}^*$ is

$$J_f^{-1} dB_{\ell_f}(u) = a(\beta)(c' z_u) du + J_f^{-1} d[Z_{\ell_f}(u) - u Z_{\ell_f}(1)], \quad (3.5.14)$$

where $B^{Z_{\ell_f}}$ is the Brownian bridge of Z_{ℓ_f} . Here (3.5.6) is omitted for the reason that it is implied by (3.5.7) and the correlation $\text{Cov}(W_\varepsilon, W_{\ell_f}) = I_2$. Moreover, (3.5.7) is identical to (3.5.14) since $\mu_z = 0$. Following the same argument as in the Gaussian case, we define the counterparts of \mathcal{S}_ϕ and \mathcal{T}_ϕ for this general case as follows

$$\mathcal{S}_f = D_{zz}^{-1/2} \int_0^1 z_u d[B_{\ell_f}(u)' J_f^{-1} b_0] \cdot (b'_0 J_f^{-1} b_0)^{-1/2}, \quad (3.5.15)$$

$$\mathcal{T}_f = D_{zz}^{-1/2} \int_0^1 z_u d[B_{\ell_f}(u)' a_0] \cdot (a'_0 J_f a_0)^{-1/2}. \quad (3.5.16)$$

Here \mathcal{S}_f and \mathcal{T}_f are both normally distributed, with respective means $D_{zz}^{1/2} c(\beta - \beta_0)(b'_0 J_f^{-1} b_0)^{-1/2}$ and $D_{zz}^{1/2} c[a(\beta) J_f a_0](a'_0 J_f a_0)^{-1/2}$, and identity covariance matrices I_k . Then, the corresponding AR, LM, and CLR type test statistics in this case are $\mathcal{AR}_f = \mathcal{S}'_f \mathcal{S}_f$, $\mathcal{LM}_f = (\mathcal{S}'_f \mathcal{T}_f)^2 / \mathcal{T}'_f \mathcal{T}_f$, and $\mathcal{CLR}_f = 0.5 \left(\mathcal{S}'_f \mathcal{S}_f - \mathcal{T}'_f \mathcal{T}_f + \sqrt{(\mathcal{S}'_f \mathcal{S}_f - \mathcal{T}'_f \mathcal{T}_f)^2 + 4(\mathcal{S}'_f \mathcal{T}_f)^2} \right)$.

Unsurprisingly, when $\beta_0 = 0$, these test statistics are the same as those introduced in Cattaneo, Crump, and Jansson (2012) in the limit. Moreover, the authors also propose the \hat{f} -based versions of the AR, LM, and CLR tests in their paper. Thus, following the methodology proposed in Section 3.3, we propose three rank-based counterparts in next section. These rank-based tests are different from the ones proposed by Andrews, and Soares (2007) and Andrews, and Marmer (2008) since we use the componentwise ranks of both series of innovations.

3.5.4 Rank-based Efficient Tests

The rank-based test statistics are constructed by the second method proposed in Section 3.3. First, recall the relationships in (3.3.2) and (3.3.3), we replace B_{ℓ_f} by $\Sigma^{-\frac{1}{2}} B_{\ell_{f\nu}}$ and J_f by $\Sigma^{-\frac{1}{2}} J_{f\nu} \Sigma^{-\frac{1}{2}'} in the “oracle” statistics \mathcal{S}_f and \mathcal{T}_f as follows$

$$\begin{aligned}\mathcal{S}_f &= D_{zz}^{-1/2} \int_0^1 z^{(T)}(u-) d \left[B_{\ell_{f\nu}}(u)' J_{f\nu}^{-1} \Sigma^{\frac{1}{2}} b_0 \right] \cdot \left(b_0' \Sigma^{\frac{1}{2}'} J_{f\nu}^{-1} \Sigma^{\frac{1}{2}} b_0 \right)^{-1/2}, \\ \mathcal{T}_f &= D_{zz}^{-1/2} \int_0^1 z^{(T)}(u-) d \left[B_{\ell_{f\nu}}(u)' \Sigma^{-\frac{1}{2}'} a_0 \right] \cdot \left(a_0' \Sigma^{-\frac{1}{2}} J_{f\nu} \Sigma^{-\frac{1}{2}'} a_0 \right)^{-1/2}.\end{aligned}$$

As the second step, for the finite-sample feasible versions, we replace $B_{\ell_{f\nu}}$ by $\hat{B}_{\ell_g}^{(T)}$, $J_{f\nu}$ by J_p , and Σ by $\hat{\Sigma}$. The resulting counterparts are given by

$$\begin{aligned}\hat{\mathcal{S}}_g &= \hat{D}_{zz}^{-1/2} \int_0^1 z^{(T)}(u-) d \left[\hat{B}_{\ell_g}^{(T)}(u)' J_p^{-1} \hat{\Sigma}^{\frac{1}{2}} b_0 \right] \cdot \left(b_0' \hat{\Sigma}^{\frac{1}{2}'} J_p^{-1} \hat{\Sigma}^{\frac{1}{2}} b_0 \right)^{-1/2}, \\ \hat{\mathcal{T}}_g &= \hat{D}_{zz}^{-1/2} \int_0^1 z^{(T)}(u-) d \left[\hat{B}_{\ell_g}^{(T)}(u)' \hat{\Sigma}^{-\frac{1}{2}'} a_0 \right] \cdot \left(a_0' \hat{\Sigma}^{-\frac{1}{2}} J_p \hat{\Sigma}^{-\frac{1}{2}'} a_0 \right)^{-1/2}.\end{aligned}$$

where $\hat{D}_{zz} = Z_T' Z_T$. Denote by \mathcal{S}_g and \mathcal{T}_g the limits of $\hat{\mathcal{S}}_g^{(T)}$ and $\hat{\mathcal{T}}_g^{(T)}$, which immediately follow from the existing weak convergence results, i.e.,

$$\begin{aligned}\mathcal{S}_g &= D_{zz}^{-1/2} \int_0^1 z_u d \left[B_{\ell_g}(u)' J_p^{-1} \Sigma^{\frac{1}{2}} b_0 \right] \cdot \left(b_0' \Sigma^{\frac{1}{2}'} J_p^{-1} \Sigma^{\frac{1}{2}} b_0 \right)^{-1/2}, \\ \mathcal{T}_g &= D_{zz}^{-1/2} \int_0^1 z_u d \left[B_{\ell_g}(u)' \Sigma^{-\frac{1}{2}'} a_0 \right] \cdot \left(a_0' \Sigma^{-\frac{1}{2}} J_p \Sigma^{-\frac{1}{2}'} a_0 \right)^{-1/2}.\end{aligned}$$

Since z_u is a k -dimensional deterministic process, hence, \mathcal{S}_g and \mathcal{T}_g are normally distributed with zero mean and identity covariance matrix I_k . The

corresponding rank-based AR, LM and CLR test statistics are given by

$$\begin{aligned} AR_g &= \widehat{S}_g' \widehat{S}_g, \\ LM_g &= (\widehat{S}_g' \widehat{T}_g)^2 / (\widehat{T}_g' \widehat{T}_g), \\ CLR_g &= \frac{1}{2} \left[\widehat{S}_g' \widehat{S}_g - \widehat{T}_g' \widehat{T}_g + \sqrt{(\widehat{S}_g' \widehat{S}_g - \widehat{T}_g' \widehat{T}_g)^2 + 4(\widehat{S}_g' \widehat{T}_g)^2} \right]. \end{aligned}$$

Under H_0 , the statistics AR_g , LM_g and CLR_g share the same limits of AR_ϕ , LM_ϕ and CLR_ϕ , as \mathcal{S}_g and \mathcal{T}_g share the same limits of \mathcal{S}_ϕ and \mathcal{T}_ϕ . Thus, We employ the same critical values of the traditional AR, LM and CLR tests for AR_g , LM_g and CLR_g , respectively.

3.5.5 Monte Carlo Simulations

In this section, we will study these testing procedures with some Monte Carlo simulations. First, we will use the large-sample results to show that, (i) these tests are asymptotically valid, and (ii) with some non-Gaussian density f , our rank-based versions are more powerful than the original AR, LM and CLR tests. Then, we will show that our tests work well in small sample cases.

Monte Carlo setup

In this Monte Carlo study, we consider the reduced model in (3.5.2) with no exogenous variables x_t . The number of instrument variables k equals four. The local correlation parameter of the endogenous variable and instruments c is fixed to be a all-ones vector with dimension k . We simulate 20,000 independent replications from this data generating process with sample sizes $T = 2500$ for the large sample case and $T = 250$ for the small sample case. Still, ρ denotes the correlation between the two innovation series, and the significance levels are chosen to be 5% for all cases.

Large-sample case

In Figure 3.4, we plot the power functions for the case where the true joint density f is multivariate t_3 , and the marginal reference densities g_1 and g_2

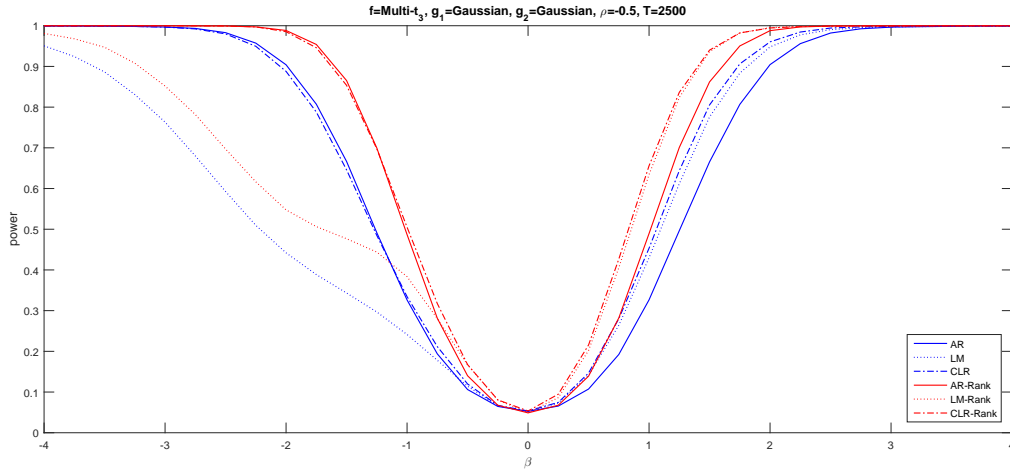


Figure 3.4: Power functions of the AR, LM, and CLR tests (the blue ones), and power functions of the Rank-based AR, LM, and CLR tests (the red ones).

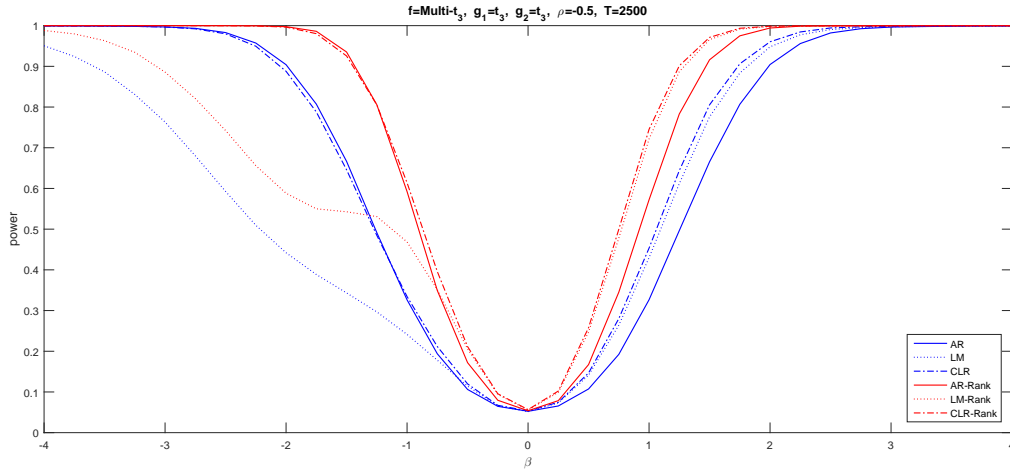


Figure 3.5: Power functions of the AR, LM, and CLR tests (the blue ones), and power functions of the Rank-based AR, LM, and CLR tests (the red ones).

are Gaussian. As we can see, the rank-based AR, LM, and CLR tests are more powerful than AR, LM, and CLR tests, respectively. The power gain is substantial, e.g., for the alternative point $\beta = 1$, the rejection rate is

about 32% for the AR test, while it is about 49% for the rank-based AR test. Similarly, for LM type tests, the rejection rate increases from 43% to 63%; for CLR tests, it increases from 45% to 66%.

In Figure 3.5, we keep multi- t_3 being the true density f , while changes the marginal reference densities g_1 and g_2 to univariate t_3 . For the same alternative point $\beta = 1$, the increases in the rejection rate for these three cases is even higher — by around 30% more for each case.

Small-sample case

Figure 3.6 can be regarded as the small-sample counterpart of Figure 3.4 with $T = 250$ and everything else left unchanged. These plots show that all tests control their sizes decently, while the rank-based tests remain more powerful than the original ones.

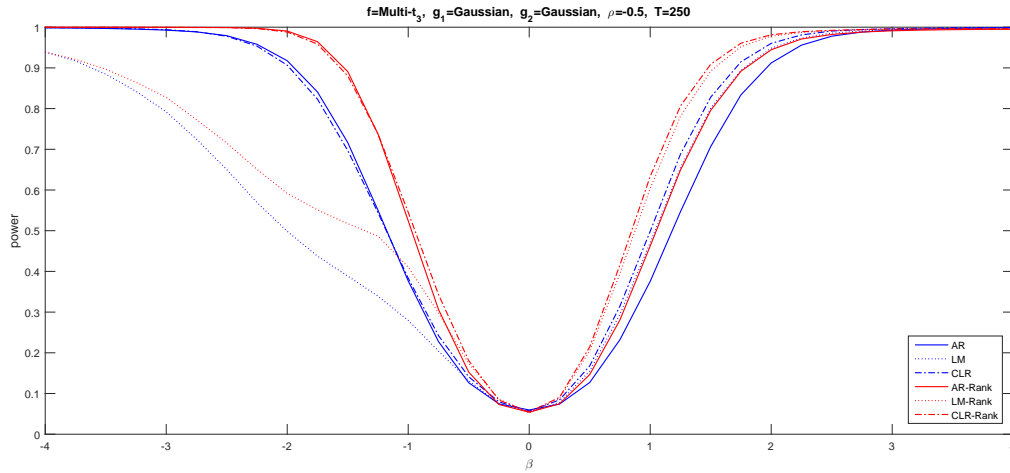


Figure 3.6: Power functions of the AR, LM, and CLR tests (the blue ones), and power functions of the Rank-based AR, LM, and CLR tests (the red ones).

3.6 Conclusion

This paper extends the branch of the semiparametric literature, where the underlying innovation density is treated as the nonparametric nuisance part, to nonstandard econometric problems. For this purpose, we provide a new semiparametric approach to derive the semiparametric power envelope of all asymptotically invariant tests. This approach is general in the sense that it works for all models in which the associated limit experiments admit the locally asymptotically Brownian functional form. Furthermore, the obtained structure of these semiparametric power envelopes, which can be written into two expressions, shows that efficient test statistics can be constructed either based on a nonparametric estimates of the unknown density function, or based on rank statistics with an arbitrarily chosen reference density.

The key step of our new approach is the structural version of the limit experiment: It clearly shows that the invariance restriction is the one we should employ to eliminate the nuisance parameter. Moreover, it also suggests solution for dealing with other existing nuisance parameters rather than the density, at least under the null hypothesis. We exploit the latter property in the weak instrument application.

We focus on rank-based semiparametric inference in the present paper. For the cases of multidimensional innovations, we propose a new scheme for constructing test statistics using componentwise ranks and marginal reference densities. In this way, on the one hand, we only use information from the marginal densities of a joint density; while, on the other hand, no additional assumption on the density function is needed. Interestingly, the simulation result of the cointegration case indicates that most, if not all, information for statistical inference efficiency consists in the marginal densities rather than the copula structure of a joint innovation density.

These new methods are applied to two nonstandard econometric problems: inference for cointegration rank testing and inference for linear regression with weak instruments. For both cases, we develop the semiparametric power envelopes and propose the rank-based testing procedures. Simula-

tion results show that our newly proposed rank-based tests are of significant more power than the commonly-used tests when the innovation density is not Gaussian, and as well as the latter when the density is Gaussian.

For future research, we can apply this new approach to more nonstandard econometric problems, for example, predictive regressions with persistent predictors; or extend the approach to more types of semiparametric models. Another interesting and meaningful direction is using some numerical methods to make rank-based semiparametric inferences robust to a wider range of data generating processes.

3.7 Appendix: Proofs

Proof of Proposition 3.2.1.

For notational convenience we drop the superscript “(T)” in the following and thus write f_η instead of $f_\eta^{(T)}$. First of all, f_η integrates to 1 for the reason that $\int_{\mathbb{R}^p} b_k(e)f(e)de = 0$ and η has finite support. Part (a): absolutely continuous property of f_η comes directly from this property of f and twice continuously differentiability of b_k . Part (c): mean zero property $\int_{\mathbb{R}^p} ef_\eta(e)de = 0$ is implied by $\int_{\mathbb{R}^p} eb_k(e)f(e)de = 0$. There exists $C_1 < \infty$ such that we have, for all T , $\|1 + T^{-1/2} \sum_{k=1}^\infty \eta_k b_k\| \leq C_1$. Together with $\int_{\mathbb{R}^p} e^2 f(e)de < \infty$, we have the finite-variance property $\int_{\mathbb{R}^p} e^2 f_\eta(e)de < \infty$. Part (b): the almost-everywhere derivative of f_η is given by

$$\dot{f}_\eta(e) = \dot{f}(e) \left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^\infty \eta_k b_k(e) \right) + f(e) \frac{1}{\sqrt{T}} \sum_{k=1}^\infty \eta_k b'_k(e).$$

Similarly, there exists $C_2 < \infty$ such that we have, for all T , $\|T^{-1/2} \sum_{k=1}^\infty \eta_k \dot{b}_k\| \leq C_2$. Moreover, there exists C_3 , for $T \geq T'$, $\|(1 + T^{-1/2} \sum_{k=1}^T \eta_k b_k)^{-1}\|_\infty^2 \leq C_3$. Then, we have

$$\int_{\mathbb{R}^p} \left(-\frac{\dot{f}_\eta(e)}{f_\eta(e)} \right)^2 f_\eta(e)de \leq 2C_1 J_f + 2C_2 C_3 < \infty,$$

which completes the proof. \square

Proof of Theorem 3.2.1.

This results follow a direct application of Girsanov's Theorem on the limit likelihood ratio (Radon-Nikodym derivative) in equation (3.2.10). Detailed algebra calculation is omitted. \square

Proof of Theorem 3.2.2.

Assume, throughout this proof, that $u \in [0, 1]$ and $\{i, j\} \in \mathcal{S} = \{\{0, \eta\}, \{m, 0\}, \{m, \eta\}\}$.

Define the maximal invariant function $M_{i,j}$ by $\mathcal{M}_{i,j} = \sigma(M_{i,j}(W_\varepsilon, W_b))$.

Specifically,

$$\begin{aligned} M_{0,\eta}(W_\varepsilon, W_b) &= (W_\varepsilon, B_b), \\ M_{m,0}(W_\varepsilon, W_b) &= (J_{bf}W_\varepsilon - W_b, B_b), \\ M_{m,\eta}(W_\varepsilon, W_b) &= (B_\varepsilon, B_b). \end{aligned}$$

The logic of the present proof originates from the definition of “*maximal invariant*” in Section 6.2 of Lehmann and Romano (2005): $\mathcal{M}_{i,j}$ is called maximal invariant with respect to $\mathfrak{G}_{i,j}$ (i) if it is invariant; and (ii) if an equality of $M_{i,j}(W_\varepsilon, W_b)$ and $M_{i,j}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$ implies that (W_ε, W_b) can be translated to $(\widetilde{W}_\varepsilon, \widetilde{W}_b)$ by some transformation $\mathbf{g}_{i,j}$ belonging to $\mathfrak{G}_{i,j}$. Condition (i) is trivially satisfied as discussed in Section 3.2.4, and a case-by-case proof for condition (ii) goes as follows:

- (a) $\mathcal{M}_{0,\eta} = \sigma(W_\varepsilon(u), B_b(u))$ is maximal invariant w.r.t the transformation group $\mathfrak{G}_{0,\eta}$: suppose $M_{0,\eta}(W_\varepsilon, W_b) = M_{0,\eta}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$, that is, explicitly

$$W_\varepsilon(u) = \widetilde{W}_\varepsilon(u), \quad \text{and} \quad B_b(u) = \widetilde{B}_b(u).$$

It implies, for $c_\eta = W_b(1) - \widetilde{W}_b(1)$, that

$$W_\varepsilon(u) - \widetilde{W}_\varepsilon(u) = 0, \quad \text{and} \quad W_b(u) - \widetilde{W}_b(u) = c_\eta u.$$

This shows that the equality of $M_{0,\eta}(W_\varepsilon, W_b)$ and $M_{0,\eta}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$ implies the former can be translated to the latter by a transformation $\mathbf{g}_{0,\eta} \in \mathfrak{G}_{0,\eta}$ with c_η defined above, which completes the proof.

- (b) $\mathcal{M}_{m,0} = \sigma(J_{bf}W_\varepsilon(u) - W_b(u), B_{b_k}(u))$ is maximal invariant w.r.t the transformation group $\mathfrak{G}_{m,0}$: suppose $M_{m,0}(W_\varepsilon, W_b) = M_{m,0}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$, that is, explicitly

$$J_{bf}W_\varepsilon(u) - W_b(u) = J_{bf}\widetilde{W}_\varepsilon(u) - \widetilde{W}_b(u), \quad \text{and} \quad B_b(u) = \widetilde{B}_b(u).$$

This implies, for $c_m = J_{bf}^{-1}(W_b(1) - \widetilde{W}_b(1))$, that

$$W_\varepsilon(u) - \widetilde{W}_\varepsilon(u) = c_m u, \quad \text{and} \quad W_b(u) - \widetilde{W}_b(u) = J_{bf}c_m u.$$

This shows that the equality of $M_{m,0}(W_\varepsilon, W_b)$ and $M_{m,0}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$ implies the former can be translated to the latter by a transformation $\mathfrak{g}_{m,0} \in \mathfrak{G}_{m,0}$ with c_m defined above, which completes the proof.

- (c) $\mathcal{M}_{m,\eta} = \sigma(B_\varepsilon(u), B_{b_k}(u))$ is maximal invariant w.r.t the transformation group $\mathfrak{G}_{m,\eta}$: suppose $M_{m,\eta}(W_\varepsilon, W_b) = M_{m,\eta}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$, that is, explicitly

$$B_\varepsilon(u) = \widetilde{B}_\varepsilon(u), \quad \text{and} \quad B_b(u) = \widetilde{B}_b(u).$$

It implies, for $c_m = W_\varepsilon(1) - \widetilde{W}_\varepsilon(1)$ and $c_\eta = W_b(1) - \widetilde{W}_b(1) - J_{bf}c_m$, that

$$W_\varepsilon(u) - \widetilde{W}_\varepsilon(u) = c_m u, \quad \text{and} \quad W_b(u) - \widetilde{W}_b(u) = (J_{bf}c_m + c_\eta) u.$$

This shows that the equality of $M_{m,\eta}(W_\varepsilon, W_b)$ and $M_{m,\eta}(\widetilde{W}_\varepsilon, \widetilde{W}_b)$ implies the former can be translated to the latter by a transformation $\mathfrak{g}_{m,\eta} \in \mathfrak{G}_{m,\eta}$ with c_m and c_η defined above, which completes the proof.

□

Proof of Theorem 3.3.1.

Proof of Part (a): This part of the proof follows from Theorem 13.5 in Van der Vaart (2000). Firstly, I take $i = 1$ as example. In the notation of Van der Vaart (2000), let $i = t$, $N = T$, $c_{Ni} = \mathbb{1}\{t \leq uT\}$, and $a_{N,R_{N,i}} = \ell_{g_1}[G_1^{-1}(R_{1,t}/(T+1))]$. Moreover, we have $\bar{a}_N \Rightarrow 0$ since $\int_0^1 \ell_{g_1}(G_1^{-1}(u))du = 0$, and $\bar{c}_N = \lfloor uT \rfloor / T \Rightarrow u$. Express the conclusion of Theorem 13.5,

$T_N = \tilde{T}_N + o_{p_0}(1)$, in the notation of the present paper, with multiplying $1/\sqrt{T}$ on both sides, we have

$$B_{\ell_{g_1}}^{(T)}(u) = W_{\ell_{g_1}}^{(T)}(u) - uW_{\ell_{g_1}}^{(T)}(1) + o_{p_0}(1) \quad (3.7.1)$$

where

$$B_{\ell_{g_1}}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_{g_1} \left[G_1^{-1} \left(\frac{R_{1,t}}{T+1} \right) \right],$$

$$W_{\ell_{g_1}}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_{g_1} \left[G_1^{-1} (F_{\nu,1}(\nu_1)) \right].$$

By Donsker's Theorem, $W_{\ell_{g_1}}^{(T)}$ converges to a Brownian motion $W_{\ell_{g_1}}$. As a result, the right hand side of (3.7.1) converges to a Brownian bridge, and so does $B_{\ell_{g_1}}^{(T)}$. Similar results hold also for $i = 2, \dots, p$. Thus, $B_{\ell_g}^{(T)}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \ell_g(R_t)$ converges to a multidimensional Brownian bridge B_{ℓ_g} .

Proof of Part (b): Again, taking $i = 1$ as example. Denote the i -th element of $W_{\ell_{f_\nu}}$ by $W_{\ell_{f_\nu,i}}$. The covariance of $W_{\ell_{g_1}}$ and $W_{\ell_{f_\nu,1}}$ is given by

$$\begin{aligned} J_{f_\nu g,1} &= \int_0^1 \cdots \int_0^1 -\frac{\dot{f}_{\nu,1}}{f_\nu}(\nu_t) \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] f_\nu(\nu_t) d\nu_{1,t} \cdots d\nu_{p,t} \\ &= \int_0^1 \cdots \int_0^1 -\dot{f}_{\nu,1}(\nu_t) \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \cdots d\nu_{p,t} \\ &= \int_0^1 \left[\int_0^1 \cdots \int_0^1 -\dot{f}_{\nu,1}(\nu_t) d\nu_{2,t} \cdots d\nu_{p,t} \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\ &= \int_0^1 \left[-\frac{\partial}{\partial \nu_{1,t}} \int_0^1 \cdots \int_0^1 f(\nu_t) d\nu_{2,t} \cdots d\nu_{p,t} \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\ &= \int_0^1 \left[-\frac{\partial}{\partial \nu_{1,t}} F_{\nu,1}(\nu_{1,t}) \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\ &= \int_0^1 \ell_{f_\nu,1}(\nu_{1,t}) \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] f_{\nu,1}(\nu_{1,t}) d\nu_{1,t} \\ &= \int_0^1 \ell_{f_\nu,1} [F_{\nu,1}^{-1}(u)] \ell_{g_1} [G_1^{-1}(u)] du, \end{aligned}$$

where the third equality comes from Leibniz's integral rule. The covariance of $W_{\ell_{g_1}}$ and $W_{\ell_{f_\nu,j}}$, $j = 2, \dots, p$, is given by

$$\int_0^1 \cdots \int_0^1 -\frac{\dot{f}_{\nu,j}}{f_\nu}(\nu_t) \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] f_\nu(\nu_t) d\nu_{1,t} \cdots d\nu_{p,t}$$

$$\begin{aligned}
&= \int_0^1 \cdots \int_0^1 -\dot{f}_{\nu,j}(\nu_t) \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \cdots d\nu_{p,t} \\
&= \int_0^1 \left[\int_0^1 \cdots \int_0^1 -\dot{f}_{\nu,j}(\nu_t) d\nu_{2,t} \cdots d\nu_{p,t} \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\
&= \int_0^1 \left[-\frac{\partial}{\partial \nu_{j,t}} \int_0^1 \cdots \int_0^1 f_{\nu}(\nu_t) d\nu_{2,t} \cdots d\nu_{p,t} \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\
&= \int_0^1 \left[-\frac{\partial}{\partial \nu_{j,t}} F_{\nu,1}(\nu_{1,t}) \right] \ell_{g_1} [G_1^{-1}(F_{\nu,1}(\nu_{1,t}))] d\nu_{1,t} \\
&= 0,
\end{aligned}$$

where the last equality comes from that fact that $\frac{\partial}{\partial \nu_{j,t}} F_{\nu,1}(\nu_{1,t}) = 0$ for $j \neq 1$. Repeat the calculation for $i = 2, \dots, p$, we have the covariance matrix of $W_{\ell_g} = [W_{\ell_{g_1}}(1), \dots, W_{\ell_{g_p}}(1)]'$ and $W_{\ell_f}(1)$ is $J_{fg} = \text{diag}(J_{fg,1}, \dots, J_{fg,p})$.

Proof of Part (c): This proof follows the result of Chernoff and Savage (1958). \square

Proof of Corollary 3.3.2.

Equation (3.3.8) is already there. To get equation (3.3.9), let us firstly decompose W_{ℓ_g} by $W_{\ell_g} = AW_{\varepsilon} + BW_{\ell_f} + W_{\perp}$, where W_{\perp} is independent of W_{ε} and W_{ℓ_f} , and A, B are some matrices in $\mathbb{R}^{p \times p}$. Then we have

$$\begin{aligned}
J_{gf} &= \text{Cov}(W_{\ell_g}(1), W_{\ell_f}(1)) \\
&= \text{Cov}(AW_{\varepsilon}(1) + BW_{\ell_f}(1) + W_{\perp}(1), W_{\ell_f}(1)) \\
&= A + BJ_f.
\end{aligned}$$

Combining $dW_{\ell_g} = AdW_{\varepsilon} + BdW_{\ell_f} + W_{\perp}$ and $\mathcal{E}^*(f)$ addressed in Remark 3.2.1, we have

$$\begin{aligned}
dW_{\ell_g}(u) &= A[M(u, h)du + mdu + dZ_{\varepsilon}(u)] + B[J_f M(u, h)du + J_f mdu + J_{fb}\eta du + dZ_{\ell_f}(u)] \\
&= (A + BJ_f)M(u, h)du + [(A + BJ_f)m + BJ_{fb}\eta]du + dZ_{\ell_g}(u) \\
&= J_{gf}M(u, h)du + [J_{gf}m + BJ_{fb}\eta]du + dZ_{\ell_g}(u),
\end{aligned}$$

and then take the bridge

$$dB_{\ell_g}(u) = J_{gf}[M(u, h) - M(1, h)]du + dB^{Z_{\ell_g}(u)},$$

which completes the proof. \square

Proof of Proposition 3.4.1.

Proof of Part (a): The log-likelihood ratio is

$$\begin{aligned} L^{(T)}(C, \eta) &= \sum_{t=2}^T \log \frac{f(\Delta y_t - \frac{C}{T} y_{t-1}) \left(1 + \frac{1}{\sqrt{T}} \eta' b(\Delta y_t - \frac{C}{T} y_{t-1})\right)}{f(\Delta y_t)} \\ &= \sum_{t=2}^T \log \frac{f(\Delta y_t - \frac{C}{T} y_{t-1})}{f(\Delta y_t)} + \sum_{t=2}^T \log \left(1 + \frac{1}{\sqrt{T}} \eta' b(\Delta y_t - \frac{C}{T} y_{t-1})\right). \end{aligned}$$

(i) For the first term in the above expression, we have

$$\begin{aligned} &\sum_{t=2}^T \log \frac{f(\Delta y_t - \frac{C}{T} y_{t-1})}{f(\Delta y_t)} \\ &= \frac{1}{T} \sum_{t=2}^T (C y_{t-1})' \ell_f(\Delta y_t) - \frac{1}{2T^2} \sum_{t=2}^T (C y_{t-1})' J_f C y_{t-1} + o_{p_0}(1). \end{aligned} \quad (3.7.2)$$

A detailed proof for this part can be found in Hallin, Van den Akker, and Werker (2016) in a simpler case, where, with their notations, $\mu = 0$, $\alpha = \beta = 0$, $\Gamma = 0$ and $D = C$.

(ii) For the second term, using the assumption on b_k that it is bounded and two times continuously differentiable with bounded derivatives for all k , we have

$$\begin{aligned} &\sum_{t=2}^T \log \left(1 + \frac{1}{\sqrt{T}} \eta' b(\Delta y_t - \frac{C}{T} y_{t-1})\right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta' b(\Delta y_t) - \frac{1}{T^{3/2}} \sum_{t=1}^T \eta' \dot{b}(\Delta y_t) C y_{t-1} + \frac{1}{2} \eta' \eta + o_{p_0}(1), \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta' b(\Delta y_t) - \frac{1}{T^{3/2}} \sum_{t=1}^T \eta' J_{bf} C y_{t-1} + \frac{1}{2} \eta' \eta + o_{p_0}(1). \end{aligned} \quad (3.7.3)$$

The first equality comes directly from a Taylor-series expansion calculation on the log form and function b . The second equality comes from equalities $\frac{1}{T^{3/2}} \sum_{t=1}^T \eta' \dot{b}(\Delta y_t) C y_{t-1} = \frac{1}{T^{3/2}} \sum_{t=1}^T \eta' E(\dot{b}(e)) C y_{t-1} + o_{p_0}(1)$ and $E(\dot{b}(e)) = \int_{\mathbb{R}^p} \dot{b}(e) f(e) de = b(e) f(e)|_{-\infty}^{+\infty} - \int_{\mathbb{R}^p} b(e) f'(e) de = 0 + \int_{\mathbb{R}^p} b(e) \left(-\frac{f'(e)}{f(e)}\right) f(e) de = \int_{\mathbb{R}^p} b(e) \ell_f(e) f(e) de = J_{bf}$. Following Taylor's theorem, higher order expansions belong to $o_{p_0}(1)$ term.

Combining equation (3.7.2) and equation (3.7.3) completes the proof for Part (a).

Proof of Part (b): This part of proofs follows from Donsker's theorem and Chan and Wei (1988, Theorem 2.4).

Proof of Part (c): This part completes by verifying the Novikov condition, which follows from an application of (Karatzas and Shreve, 1991, Corollary 3.5.16). \square

Proof of Proposition 3.5.1.

Proofs for Part (b) and Part (c) uses the same approach as corresponding parts in the proof for Proposition 3.4.1. For Part (a), the log-likelihood ratio is give by

$$\begin{aligned} L^{(T)}(\beta, c, \eta) &= \sum_{t=2}^T \log \frac{f(y_{1t} - \beta \frac{z'_t c}{\sqrt{T}}, y_{2t} - \frac{z'_t c}{\sqrt{T}}) \left(1 + \frac{1}{\sqrt{T}} \eta' b \left(y_{1t} - \beta \frac{z'_t c}{\sqrt{T}}, y_{2t} - \frac{z'_t c}{\sqrt{T}}\right)\right)}{f(y_{1t}, y_{2t})} \\ &= \sum_{t=2}^T \log \frac{f(y_{1t} - \beta \frac{z'_t c}{\sqrt{T}}, y_{2t} - \frac{z'_t c}{\sqrt{T}})}{f(y_{1t}, y_{2t})} \\ &\quad + \sum_{t=2}^T \log \left(1 + \frac{1}{\sqrt{T}} \eta' b \left(y_{1t} - \beta \frac{z'_t c}{\sqrt{T}}, y_{2t} - \frac{z'_t c}{\sqrt{T}}\right)\right). \end{aligned}$$

(i) For the first term in the above expression, we have

$$\begin{aligned} &\sum_{t=2}^T \log \frac{f(y_{1t} - \beta \frac{z'_t c}{\sqrt{T}}, y_{2t} - \frac{z'_t c}{\sqrt{T}})}{f(y_{1t}, y_{2t})} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T a(\beta)'(c' z_t) \ell_f(y_{1t}, y_{2t}) - \frac{1}{2} a(\beta)'(c' D_{zz} c) J_f a(\beta) + o_{p_0}(1). \end{aligned}$$

A detailed proof for this part can be found in Cattaneo, Crump, and Jansson (2012).

(ii) The second term follows by the same argument based on Taylor's theorem in the corresponding part of proof for Proposition 3.4.1. \square

3.8 Appendix: Conditional Expectation Calculations

3.8.1 Linear Regression Case

Both m and η are unknown nuisance parameters

Recall the limit log-likelihood ratio is

$$\mathcal{L}(h, m, \eta) = \Delta(h, m, \eta) - \frac{1}{2} \mathcal{Q}(h, m, \eta) \quad (3.8.1)$$

where

$$\begin{aligned} \Delta(h, m, \eta) &= \int_0^1 (h'x_u + m) dW_{\ell_f}(u) + \eta'W_b(1), \\ \mathcal{Q}(h, m, \eta) &= J_f h' D_{xx} h + m^2 J_f + \eta' \eta + 2h' \mu_x J_{fb} \eta + 2h' \mu_x J_f m + 2m J_{fb} \eta. \end{aligned}$$

Split Δ into two parts: $\Delta = \Delta_1^I + \Delta_1^{II}$, where processes Δ_u^I and Δ_u^{II} are defined by

$$\begin{aligned} \Delta_u^I &= \int_0^u (h'x_s + m) dB_{\ell_f}(s), \\ \Delta_u^{II} &= (h' \mu_x + m) \int_0^u dW_{\ell_f}(s) + \eta' \int_0^u dW_b(s). \end{aligned}$$

Hence the quadratic part \mathcal{Q} is nothing more but the quadratic variance of the process $\Delta_u^I + \Delta_u^{II}$ at time 1. Recall that x_u is a known deterministic process and $\mathcal{M}_{m,\eta} = \sigma(B_\varepsilon(u), B_b(u), u \in [0, 1])$, it is obvious that Δ_1^I is $\mathcal{M}_{m,\eta}$ -measurable since $B_{\ell_f} = \Sigma^{-1}B_\varepsilon + J_{fb}B_b$. Δ_1^{II} is independent of $\mathcal{M}_{m,\eta}$ since Brownian motion at time 1 is independent of its associated Brownian bridge. Subsequently, Δ_1^I is independent of Δ_1^{II} . Let “ $\langle \cdot \rangle_u$ ” denote taking the quadratic variation at time u . Then \mathcal{L} can be rewrite as

$$\begin{aligned} \mathcal{L} &= \Delta_1^I + \Delta_1^{II} - 0.5 \langle \Delta^I + \Delta^{II} \rangle_1 \\ &= \Delta_1^I + \Delta_1^{II} - 0.5 \langle \Delta^I \rangle_1 - 0.5 \langle \Delta^{II} \rangle_1. \end{aligned}$$

Proceeding the calculation in equation (3.2.19) for $\mathcal{M}_{m,\eta}$, we have

$$\mathcal{L}_{\mathcal{M}_{m,\eta}}(h) = \log E_0 [\exp(\mathcal{L}(h, m, \eta)) | \mathcal{M}_{m,\eta}]$$

$$\begin{aligned}
&= \log E_0 [\exp(\Delta_1^I + \Delta_1^{II} - 0.5 \langle \Delta^I \rangle_1 - 0.5 \langle \Delta^{II} \rangle_1) | \mathcal{M}_{m,\eta}] \\
&= \log \exp(\Delta_1^I - 0.5 \langle \Delta^I \rangle_1) E_0 [\exp(\Delta_1^{II} - 0.5 \langle \Delta^{II} \rangle_1) | \mathcal{M}_{m,\eta}] \\
&= \log \exp(\Delta_1^I - 0.5 \langle \Delta^I \rangle_1) E_0 [\exp(\Delta_1^{II} - 0.5 \langle \Delta^{II} \rangle_1)] \\
&= \Delta_1^I - 0.5 \langle \Delta^I \rangle_1.
\end{aligned}$$

Note $\Delta_1^I = \int_0^1 (h'x_s + m) dB_{\ell_f}(s) = \int_0^1 h'x_s dB_{\ell_f}(s)$ and similarly for $\langle \Delta^I \rangle_1$, the constant parameter m cancels out automatically. Replacing $\mathcal{M}_{m,\eta}$ by $\mathcal{M}_{m,\eta}^*$ will lead to exactly the same result, since Δ_1^I is $\mathcal{M}_{m,\eta}^*$ -measurable and Δ_1^{II} is independent of $\mathcal{M}_{m,\eta}^*$.

m is an unknown nuisance parameter while η is known to be 0

The limit log-likelihood ratio is

$$\mathcal{L}(h, m, 0) = \Delta(h, m) - \frac{1}{2} \mathcal{Q}(h, m) \quad (3.8.2)$$

where

$$\begin{aligned}
\Delta(h, m) &= \int_0^1 (h'x_u + m) dW_{\ell_f}(u), \\
\mathcal{Q}(h, m) &= J_f h' D_{xx} h + m^2 J_f + 2h' \mu_x J_f m.
\end{aligned}$$

Split Δ into two parts: $\Delta = \Delta_1^I + \Delta_1^{II}$, where the processes Δ_u^I and Δ_u^{II} are defined by

$$\begin{aligned}
\Delta_u^I &= \int_0^u (h'x_s + m) dB_{\ell_f}(s), \\
\Delta_u^{II} &= (h' \mu_x + m) \int_0^u dW_{\ell_f}(s).
\end{aligned}$$

Recall $\mathcal{M}_{m,0}^* = \sigma(J_f W_\varepsilon(u) - W_{\ell_f}(u), B_{\ell_f}(u), u \in [0, 1])$, so Δ_1^I is $\mathcal{M}_{m,0}^*$ -measurable.

Moreover, since $\text{Cov}(J_f W_\varepsilon(u) - W_{\ell_f}(u), W_{\ell_f}(1)) = J_f u - J_f u = 0$ and $W_{\ell_f}(1)$ is independent of $B_{\ell_f}(u)$, which bring out that Δ_1^{II} is independent of $\mathcal{M}_{m,0}^*$.

Then, following the same calculations, we have

$$\begin{aligned}
\mathcal{L}_{\mathcal{M}_{m,\eta}}(h) &= \log E_0 [\exp(\mathcal{L}(h, m, 0)) | \mathcal{M}_{m,0}^*] \\
&= \Delta_1^I - 0.5 \langle \Delta^I \rangle_1,
\end{aligned}$$

which is exactly the same result as in the previous case.

3.8.2 Cointegration Case

Recall the limit log-likelihood ratio

$$\mathcal{L}(C, \eta) = \Delta(C, \eta) - \frac{1}{2}\mathcal{Q}(C, \eta) \quad (3.8.3)$$

with

$$\begin{aligned} \Delta(C, \eta) &= \int_0^1 [CW_\varepsilon(u)]' dW_{\ell_f}(u) + \eta' W_b(1), \\ \mathcal{Q}(C, \eta) &= \int_0^1 [CW_\varepsilon(u)]' J_f [CW_\varepsilon(u)] du + 2\eta' \int_0^1 J_{bf} [CW_\varepsilon(u)] du + \eta' \eta. \end{aligned}$$

Split Δ into two parts: $\Delta = \Delta_1^I + \Delta_1^{II}$, where the processes Δ_u^I and Δ_u^{II} are defined by

$$\begin{aligned} \Delta_u^I &= \int_0^u [CW_\varepsilon(s)]' dB_{\ell_f}(s) + \int_0^1 [CW_\varepsilon(s)]' ds \int_0^u \Sigma^{-1} dW_\varepsilon(s), \\ \Delta_u^{II} &= \int_0^1 [CW_\varepsilon(u)]' ds \int_0^u [dW_{\ell_f}(s) - \Sigma^{-1} dW_\varepsilon(s)] + \eta' \int_0^u dW_b(s). \end{aligned}$$

Recall $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$. It is not difficult to find that Δ_1^I is $\mathcal{M}_{0,\eta}^*$ -measurable, and Δ_1^{II} is independent of $\mathcal{M}_{0,\eta}^*$ since a Brownian motion at time 1 is independent of its associated Brownian bridge and $\text{Cov}(W_{\ell_f}(1) - \Sigma^{-1}W_\varepsilon(1), W_\varepsilon(u)) = 0$. Similarly as the conditional expectation calculation in Section 3.8.1, it has

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(C) = \log E_0 [\exp(\mathcal{L}(C, \eta)) | \mathcal{M}_{0,\eta}^*] = \Delta_1^I - 0.5 \langle \Delta^I \rangle_1$$

where $\Delta_1^I = \Delta_{\mathcal{M}_{0,\eta}^*}(C)$ and $\langle \Delta^I \rangle_1 = \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(C)$.

3.8.3 Weak Instrument Case

From the limiting log-likelihood ratio we obtain the LAN result

$$\mathcal{L}(\beta, c, \eta) = \Delta(\beta, c, \eta) - \frac{1}{2}\mathcal{Q}(\beta, c, \eta) \quad (3.8.4)$$

with

$$\Delta(\beta, c, \eta) = \int_0^1 a(\beta)'(c'z_u) dW_{\ell_f}(u) + \eta' W_b(1),$$

$$\mathcal{Q}(\beta, c, \eta) = a(\beta)'(c'D_{zz}c)J_f a(\beta) + 2a(\beta)'J_{fb}\eta \cdot c'\mu_z + \eta'\eta.$$

Split Δ into two parts: $\Delta = \Delta_1^I + \Delta_1^{II}$, where processes Δ_u^I and Δ_u^{II} are defined by

$$\begin{aligned}\Delta_u^I &= \int_0^u a(\beta)'(c'z_s)dB_{\ell_f}(s) + \int_0^1 a(\beta)'(c'z_s)ds \int_0^u dW_\varepsilon(s), \\ \Delta_u^{II} &= \int_0^1 a(\beta)'(c'z_u)du \int_0^u [dW_{\ell_f}(s) - dW_\varepsilon(s)] + \eta' \int_0^u dW_b(s).\end{aligned}$$

Recall $\mathcal{M}_{0,\eta}^* = (W_\varepsilon(u), B_{\ell_f}(u), u \in [0, 1])$, then it has that Δ_1^I is $\mathcal{M}_{0,\eta}^*$ -measurable and Δ_1^{II} is independent of $\mathcal{M}_{0,\eta}^*$ following the same reason in Section 3.8.1.

Similarly

$$\mathcal{L}_{\mathcal{M}_{0,\eta}^*}(\beta, c) = \log E_0 [\exp(\mathcal{L}(\beta, c, \eta)) | \mathcal{M}_{0,\eta}^*] = \Delta_1^I - 0.5 \langle \Delta^I \rangle_1$$

where

$$\begin{aligned}\Delta_1^I &= \Delta_{\mathcal{M}_{0,\eta}^*}(\beta, c) = a(\beta)' \int_0^1 c'z_u dB_{\ell_f}(u), \\ \langle \Delta^I \rangle_1 &= \mathcal{Q}_{\mathcal{M}_{0,\eta}^*}(\beta, c) = a(\beta)' [(c'D_{zz}c) - (c'\mu_z)^2] J_f a(\beta).\end{aligned}$$

since $\mu_z = 0$ leads to the second term of Δ_1^I , $\int_0^1 a(\beta)'(c'z_s)ds W_\varepsilon(1)$, equaling zero.

Chapter 4

Semiparametric Optimal Testing with Highly Persistent Predictors

[Based on joint work with Ramon van den Akker and Bas Werker
Semiparametric Optimal Testing with Highly Persistent Predictors.]

Abstract. Consider a bivariate regression model with a highly persistent predictor, where the joint distribution of the innovations is an infinite-dimensional nuisance parameter. Using a structural representation of the limit experiment and exploiting invariance relationships therein, we construct asymptotically invariant point-optimal tests for the regression coefficient of interest. This approach naturally leads to a family of tests based on the component-wise ranks of the innovations. Simulations show that this family of tests gains considerable power relative to existing tests when the innovations are non Gaussian.

Key words. Predictive regression, limit experiment, LABF, maximal invariant, rank statistic.

4.1 Introduction

Over the past two decades, inference for the bivariate regression model with a highly persistent regressor has been well studied under the assumption of bivariate Gaussian innovations. Several procedures have been proposed in the econometric literature, among which there are Campbell and Yogo (2005), Cavanagh, Elliott, and Stock (1995), Elliott, Müller, and Watson (2015), and Jansson and Moreira (2006). These inferences procedures are all constructed based on the assumption of Gaussian innovations. Thus, the asymptotic powers of all these inferences cannot go beyond the Gaussian power envelope.

The study of optimal semiparametric inference in the predictive regression model is complicated by the nonstandard asymptotic behavior induced by the local-to-unity asymptotics on the persistence parameter. More precisely, the associated likelihood ratios are of the Locally Asymptotically Brownian Functional (LABF) form (Jeganathan (1995)) and henceforth outside the conventional Locally Asymptotically Normality (LAN) world. As a consequence, the usual semiparametric approach based on projecting the score of the parameter of interest on the tangent space of nuisance scores is not tractable. Jansson (2008) deals with the unit root testing problem, for which the model also admits the LABF form, by guessing a least favorable direction of parametric submodels. However, this approach does not seem to be straightforward to extend to other cases. Alternatively, a new approach has been proposed recently for the problem of testing for unit roots in Zhou, van den Akker, and Werker (2016) and has been generalized to all LAN, LAMN and LABF models by Zhou (2017).¹ In the present paper we apply these techniques to the predictive regression model.

Our contribution is twofold. First, we derive the semiparametric power envelope for (asymptotically) invariant tests in case the regressor's persistence level is assumed to be known. This step is based on a structural

¹Here "LAMN" is short for Locally Asymptotically Mixed Normality (see its definition, e.g., in Jeganathan (1995)).

representation of LABF limit experiments. More precisely, Girsanov's theorem, combined with the limiting likelihood ratios for LABF experiments, leads to a structural version of the limit experiment described by stochastic differential equations (SDEs). The observations in the limit experiment correspond to the limits of partial-sum processes of the errors and score functions in the predictive regression model. In this structural representation of the limit experiment, we find that the nuisance parameters induced by the intercept and the density function of the innovations only appear in the drifts of the SDEs. This suggests an invariance restriction, in particular, by taking the (Brownian) bridges of these processes. This allows us to eliminate these nuisance parameters by invariance arguments, thus avoiding the problem of explicitly finding the least favorable submodel. The likelihood of the maximal invariant immediately leads to the semiparametric power envelope.

As a second contribution, we propose a family of semiparametric feasible tests that has some desirable properties. These tests are constructed using (asymptotically) sufficient statistics that are based on the increments of innovations, their componentwise ranks, and chosen marginal reference densities for both innovations. The ranks appear naturally as rank-based partial sum processes weakly converge to precisely the Brownian bridges that appear in the limiting maximal invariant. To eliminate the last remaining nuisance parameter, namely the regressor's persistence level, we employ the Approximate Least Favorable Distribution approach proposed by Elliott, Müller, and Watson (2015). The tests thus obtained are semiparametric in the sense that they have correct asymptotic sizes, regardless of the choices of the marginal reference densities and regardless of the true underlying innovation distributions.² Furthermore, they are more powerful than existing tests when the true innovation density is non-Gaussian. Finally, they perform as existing tests in the case that both the true and the

²Unlike the traditional QMLE methods, the reference densities need not necessarily be Gaussian.

reference densities are Gaussian.

The paper is organized as follows. In Section 4.2, we formally introduce the model and testing problem under consideration. In Section 4.3, we will develop the asymptotic power envelope for test that are (asymptotically) invariant with respect to μ and f , assuming γ is known. The development is based on the theory of limit experiments (see Le Cam (1986) and Van der Vaart (2000)) and a structural version for models of Locally Asymptotically Brownian Functional (LABF) likelihood ratios (see Zhou, van den Akker, and Werker (2016) and Zhou (2017)). In Section 4.4, we employ the Approximate Least Favorable Distribution (ALFD) approach proposed by Elliott, Müller, and Watson (2015), among several available choices in the literature, to eliminate the nuisance parameter γ . Section 4.5 reports both large and small sample performances of our tests, while Section 4.6 concludes. All proofs are gathered in the appendix.

4.2 Model

Let y_t denote a random variable observable at time t , that we wish to predict at time $t - 1$ using an observable explanatory variable x_{t-1} . We consider the predictive regression model

$$y_t = \mu + \beta x_{t-1} + \varepsilon_t^y, \quad (4.2.1)$$

$$x_t = \gamma x_{t-1} + \varepsilon_t^x, \quad (4.2.2)$$

with $x_0 = 0$.³ The parameter space is given by $\mu \in \mathbb{R}$, $\beta \in \mathbb{R}$, and $\gamma \in (-1, 1]$. We have observations available for $t = 1, \dots, T$. Observe that, in line with most of the literature, (4.2.2) does not feature an intercept. Adding such an intercept would lead to a different asymptotic analysis.

³Note that this assumption on the initial value x_0 in the present paper could be possibly relaxed, along the line of Müller and Elliott (2003), to a weaker assumption that $T^{-1/2}x_0 = o_P(1)$ under $\beta = 0$ and $\gamma = 1$. A similar conjecture for Gaussian density case can be found in Section 4 of Jansson and Moreira (2006). We keep the assumption $x_0 = 0$ for simplicity.

We assume that the innovations $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)'$ are independent and identically distributed (i.i.d.) with (bivariate) density f , satisfying the following assumption.

Assumption 4.2.1. We impose the following restrictions on the innovation density f .

- (a) $E_f(\varepsilon_t) = 0$ and $\text{Var}_f(\varepsilon_t) = \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & \sigma_x^2 \end{pmatrix}$ is a finite positive-definite matrix.
- (b) The density f is absolutely continuous with a.e. derivative $\dot{f} = \begin{pmatrix} \dot{f}_y \\ \dot{f}_x \end{pmatrix}$.
- (c) The (standardized) Fisher information for location,

$$J_f = \begin{pmatrix} J_{f_{yy}} & J_{f_{yx}} \\ J_{f_{yx}} & J_{f_{xx}} \end{pmatrix} = E_f(\ell_f \ell_f'),$$

where ℓ_f is the (standardized) location score function

$$\ell_f = \begin{pmatrix} \sigma_y \ell_{f_y} \\ \sigma_x \ell_{f_x} \end{pmatrix} = \begin{pmatrix} -\sigma_y \dot{f}_y / f \\ -\sigma_x \dot{f}_x / f \end{pmatrix},$$

is finite.⁴

- (d) $f > 0$.

□

Let \mathfrak{F} denote the set of densities satisfying Assumption 4.2.1.

The Fisher information J_f and scores ℓ_f for location are standardized in the sense that they are actually those related to ε_t^y/σ_y and ε_t^x/σ_x . As a result, ℓ_f and J_f do not depend on σ_y and σ_x . Note, however, that they both still depend on the correlation between the innovations ε_t^y and ε_t^x , i.e., they still depend on ρ .

⁴Being a Fisher information for location, it is automatically nonsingular, positive definite, see Mayer-Wolf (1990, Theorem 2.3).

We are interested in optimal tests for the (composite) null hypothesis

$$H_0 : \beta = 0, \mu \in \mathbb{R}, \gamma \in (-1, 1], f \in \mathfrak{F} \quad (4.2.3)$$

versus the one-sided alternative⁵

$$H_1 : \beta > 0, \mu \in \mathbb{R}, \gamma \in (-1, 1], f \in \mathfrak{F}. \quad (4.2.4)$$

4.2.1 Preliminaries

Following the by now standard approach in the literature, we consider the limit experiment in the sense of Hájek-Le Cam by considering local alternatives for all model parameters, that is, for both the parameter of interest β and all nuisance parameters (μ , γ , and f). For μ , β , and γ the appropriate rates are well known, see, e.g., Elliott and Stock (1994), Campbell and Yogo (2005) or Jansson and Moreira (2006). More precisely, we consider a T^{-1} -localization rate for β and γ , i.e.,

$$\beta = \beta^{(T)}(b) = \frac{b}{T} \frac{\sigma_y}{\sigma_x}, \quad \gamma = \gamma^{(T)}(c) = 1 + \frac{c}{T}, \quad (4.2.5)$$

with $b, c \in \mathbb{R}$.⁶ Observe that the local perturbations for b feature a scaling by σ_y/σ_x . This ensures that the limit experiment otherwise only depends on ρ , and not σ_y and σ_x . For the intercept parameter μ the classical $T^{-1/2}$ -rate is appropriate, i.e.,

$$\mu = \mu^{(T)}(m) = \mu_0 + \frac{m}{\sqrt{T}} \sigma_y, \quad (4.2.6)$$

where $\mu_0 \in \mathbb{R}$ denotes a fixed null-value of the intercept.

The nuisance parameter f is infinite dimensional, so it is somewhat more involved to describe its relevant local perturbations. Introduce the separable Hilbert space

$$L_2^{0,f} = L_2^{0,f}(\mathbb{R}^2, \mathcal{B}) = \left\{ h \in L_2^f(\mathbb{R}^2, \mathcal{B}) \mid E_f h(e) = 0, E_f e h(e) = 0 \right\},$$

⁵This specification is usual in the literature, two-sided tests can be considered as well.

⁶The development of the limit experiment and associated invariance structures does not require $c \leq 0$. We therefore do not impose this assumption.

where $L_2^f(\mathbb{R}^2, \mathcal{B})$ denotes, the space of Borel-measurable functions $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfying $E_f h^2(e) = \int_{\mathbb{R}^2} h^2(e) f(e) de < \infty$. The separability (of the Hilbert space) ensures the existence of a countable orthonormal basis $h_k, k \in \mathbb{N}$, such that each h_k is bounded and two times continuously differentiable with bounded derivatives. Therefore, any function $h \in L_2^{0,f}$ can be written as $h = \sum_{k=1}^{\infty} \eta_k h_k$, for some $\eta = (\eta_k)_{k \in \mathbb{N}} \in \ell_2 = \{(z_k)_{k \in \mathbb{N}} \mid \sum_{k=1}^{\infty} z_k^2 < \infty\}$. Besides the space ℓ_2 , we also need the space c_{00} which is defined as the subset of sequences with finite support, i.e.,

$$c_{00} = \left\{ (z_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{k=1}^{\infty} 1_{\{z_k \neq 0\}} < \infty \right\}. \quad (4.2.7)$$

Observe that c_{00} is a dense subspace of ℓ_2 .

We now model local perturbations to the innovation density f in the following way.

$$f_{\eta}^{(T)}(e) = f(e) \left(1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k h_k(e) \right) \text{ for all } e \in \mathbb{R}^2, \quad (4.2.8)$$

where $\eta \in c_{00}$. We thus effectively use a localization rate $T^{-1/2}$ for the bivariate density f . As both μ and f are nuisance parameters describing the distribution of the innovations, it is to be expected that the localization rates for μ and f are the same. Indeed, Proposition 4.3.1 below shows that all the above rates are appropriate in the sense that they lead to contiguous alternatives for the induced probability measure as T tends to infinity.

In order to show that the above localization of the innovation density is valid, we need to establish that $f_{\eta}^{(T)} \in \mathfrak{F}$. This is the contents of the next proposition.

Proposition 4.2.1. *Let $f \in \mathfrak{F}$ and $\eta \in c_{00}$, then there exists a finite integer \tilde{T} such that for all $T \geq \tilde{T}$ we have $f_{\eta}^{(T)} \in \mathfrak{F}$.*

The proof uses exactly the same arguments as in the proof of Proposition 3.1 in Zhou, van den Akker, and Werker (2016), but with support \mathbb{R}^2 instead of \mathbb{R} . It is therefore omitted.

Hypothesis of interest

In terms of the local parameters m , b , c and η , the hypothesis of interest becomes

$$H_0 : b = 0 \text{ versus } H_1 : b > 0, \quad (4.2.9)$$

where $m \in \mathbb{R}$, $c \in (-\infty, 0]$ and $\eta \in c_{00}$ are treated as nuisance parameters.⁷

Probability measures

We will denote by $P_{m,b,c,\eta;f}^{(T)}$ the law of $(y_1, x_1)', \dots, (y_T, x_T)'$ under the model (4.2.1)–(4.2.2), where the parameters β , γ and μ are given by (4.2.5)–(4.2.6), and the innovation density is given by (4.2.8). Formally, we define the sequence of experiments of interest by

$$\mathcal{E}^{(T)}(f) = \left(\Omega^{(T)}, \mathcal{F}^{(T)}, \left\{ P_{m,b,c,\eta;f}^{(T)} : m, b, c \in \mathbb{R}, \eta \in c_{00} \right\} \right), \quad T \in \mathbb{N},$$

where $\Omega^{(T)} := \mathbb{R}^{2 \times T}$ and $\mathcal{F}^{(T)} := \mathcal{B}(\mathbb{R}^{2 \times T})$. We will denote the expectation taken under the measure $P_{0,0,0,0;f}^{(T)}$ by $E^{(T)}$.

Let us already mention that we will introduce a collection of probability measures $\mathbb{P}_{m,b,c,\eta}$ representing the limit experiment $\mathcal{E}(f)$ in Section 4.3.1 below; see (4.3.3). We will denote the probability space associated to the limit experiment by $(\Omega, \mathcal{F}, \mathbb{P}_{m,b,c,\eta})$ and denote the expectation taken under the measure $\mathbb{P}_{0,0,0,0}$ by \mathbb{E} . That is, $P^{(T)}$ and $E^{(T)}$ refer to finite-sample distributions in the sequence of experiments, while \mathbb{P} and \mathbb{E} refer to distributions in the limit experiment.

Partial-sum processes

As a final ingredient for our analysis, we introduce some partial-sum processes that we will use throughout to link the sequence of experiments $\mathcal{E}^{(T)}(f)$ to the limit experiment $\mathcal{E}(f)$. In particular, define, with $\Delta x_t =$

⁷We use here the common approach in the literature to restrict the nuisance parameter c to $(-\infty, 0]$. We conjecture that all results remain valid, with the obvious modifications, in case one would choose the larger parameter space $c \in \mathbb{R}$.

$x_t - x_{t-1}$ and $y_t^\diamond = y_t - \mu_0$, the partial-sum processes

$$W_\varepsilon^{(T)}(s) := \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \frac{\Delta x_t}{\sigma_x}, \quad (4.2.10)$$

$$W_{\ell_{fy}}^{(T)}(s) := \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \sigma_y \ell_{fy}(y_t^\diamond, \Delta x_t), \quad (4.2.11)$$

$$W_{\ell_{fx}}^{(T)}(s) := \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \sigma_x \ell_{fx}(y_t^\diamond, \Delta x_t), \quad (4.2.12)$$

$$W_{h_k}^{(T)}(s) := \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} h_k(y_t^\diamond, \Delta x_t), \quad k \in \mathbb{N}. \quad (4.2.13)$$

Here we standardize the first three partial-sum processes by the standard deviations σ_y and σ_x in order to make their limits scale invariant. Define the infinite-dimensional process $W_h^{(T)} = (W_{h_k}^{(T)})'_{k \in \mathbb{N}}$. Under $\mathbb{P}_{0,0,0,0,f}^{(T)}$, by the Functional Central Limit Theorem (see also Lemma 4.7.1 in Appendix 4.7.1), we have

$$\begin{pmatrix} W_\varepsilon^{(T)}(s) \\ W_{\ell_{fy}}^{(T)}(s) \\ W_{\ell_{fx}}^{(T)}(s) \\ W_h^{(T)}(s) \end{pmatrix} \Rightarrow \begin{pmatrix} W_\varepsilon(s) \\ W_{\ell_{fy}}(s) \\ W_{\ell_{fx}}(s) \\ W_h(s) \end{pmatrix}, \quad s \in [0, 1], \quad (4.2.14)$$

where the Brownian motions W_ε , $W_{\ell_{fy}}$, $W_{\ell_{fx}}$ and W_h are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P}_{0,0,0,0})$.⁸ Here the (joint) weak convergence is taken in the product spaces of $D[0, 1]$ with the uniform topology.

Note that the weak convergence of $W_h^{(T)}$ to the Brownian motion W_h in (4.2.14) seems to be infinite dimensional. However, all convergences in the paper are effectively only finite dimensional precisely because we take the local parameter η to be in $c_0 0$. For the sake of convenient notation, we simply write the seemingly infinite dimensional convergence (4.2.14).

Define the column vectors $J_{fyh} = (J_{fyh_k})_{k \in \mathbb{N}}$ and $J_{fxh} = (J_{fxh_k})_{k \in \mathbb{N}}$, where $J_{fyh_k} := \mathbb{E}_f [\sigma_y \ell_{fy}(\varepsilon_t) h_k(\varepsilon_t)]$ and $J_{fxh_k} := \mathbb{E}_f [\sigma_x \ell_{fx}(\varepsilon_t) h_k(\varepsilon_t)]$. Also,

⁸The use of the notation $\mathbb{P}_{0,0,0,0}$ will become more clear when deriving the limit experiment below Proposition 4.3.1.

since we have $E_f [\varepsilon_t^x \ell_{f_y}(\varepsilon_t)] = 0$ and $E_f [\varepsilon_t^x \ell_{f_x}(\varepsilon_t)] = 1$.⁹ The behavior of these Brownian motions can be described by the covariance matrix

$$\text{Var} \begin{pmatrix} W_\varepsilon(1) \\ W_{\ell_{f_y}}(1) \\ W_{\ell_{f_x}}(1) \\ W_h(1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & J_{f_y y} & J_{f_y x} & J'_{f_y h} \\ 1 & J_{f_y x} & J_{f_x x} & J'_{f_x h} \\ 0 & J_{f_y h} & J_{f_x h} & I_\infty \end{pmatrix}, \quad (4.2.15)$$

where I_∞ denotes the infinite-dimensional identity matrix. The scalings by σ_x and σ_y introduced in (4.2.10)–(4.2.13) are such that the covariance matrix (4.2.15) does not depend on σ_x or σ_y . It still depends on ρ .

Recall that the functions h_k form an orthonormal basis for all zero-mean finite-variance functions that are orthogonal to $(\varepsilon_t^y, \varepsilon_t^x)$. In view of the covariance matrix (4.2.15), we may thus write

$$W_{\ell_{f_x}}(s) = W_\varepsilon(s) + J'_{f_x h} W_h(s), \quad (4.2.16)$$

$$W_{\ell_{f_y}}(s) = J'_{f_y h} W_h(s), \quad s \in [0, 1], \quad (4.2.17)$$

and, consequently,

$$\text{Var} [W_{\ell_{f_x}}(1)] = J_{f_x x} = 1 + J'_{f_x h} J_{f_x h}, \quad (4.2.18)$$

$$\text{Var} [W_{\ell_{f_y}}(1)] = J_{f_y y} = J'_{f_y h} J_{f_y h}, \quad (4.2.19)$$

$$\text{Cov} [W_{\ell_{f_y}}(1), W_{\ell_{f_x}}(1)] = J_{f_y x} = J'_{f_y h} J_{f_x h}. \quad (4.2.20)$$

4.3 Eliminating μ and f by invariance

We now first focus on eliminating the nuisance parameters μ and f from the testing problem outlined in Section 4.2. We will see that these can be handled using invariance arguments in the limit experiment, which we will derive in Section 4.3.1. In Section 4.4, we consider the nuisance parameter γ .

We take the following steps in this section.

⁹This is due to the facts that $E_f [\varepsilon_t^x \ell_{f_y}(\varepsilon_t)] = -\int_{\mathbb{R}^2} \varepsilon_t^x \frac{\dot{f}_y(\varepsilon_t)}{f(\varepsilon_t)} f(\varepsilon_t) d\varepsilon_t = -\int_{\mathbb{R}^2} \varepsilon_t^x \dot{f}_y(\varepsilon_t) d\varepsilon_t = -\varepsilon_t^x f(\varepsilon_t)|_{-\infty}^{+\infty} + 0 = 0$, and $E_f [\varepsilon_t^y \ell_{f_y}(\varepsilon_t)] = -\int_{\mathbb{R}^2} \varepsilon_t^y \frac{\dot{f}_y(\varepsilon_t)}{f(\varepsilon_t)} f(\varepsilon_t) d\varepsilon_t = -\int_{\mathbb{R}^2} \varepsilon_t^y \dot{f}_y(\varepsilon_t) d\varepsilon_t = -\varepsilon_t^y f(\varepsilon_t)|_{-\infty}^{+\infty} + \int_{\mathbb{R}^2} f(\varepsilon_t) d\varepsilon_t = 1$.

1. Provide a structural representation of the limit experiment (Section 4.3.1).
2. Characterize maximally invariant test statistics in this limit experiment (Section 4.3.2).
3. Obtain asymptotically point-optimal invariant tests in this limit experiment (Section 4.3.3).

These steps also show that, instead of studying invariance restrictions in the *sequence* of finite-sample experiments, we only impose them in the *limit* experiment (see Müller (2011, Section 3.2) for an associated argument). For instance, Jansson and Moreira (2006) impose, for finite T , invariance of tests with respect to the parameter μ . We only impose invariance of the tests in the limit experiment and, as a result, this invariance argument then extends to the full innovation distribution, not just the location of ε_t^y . The price to pay is that we loose exact finite-sample invariance properties. However, Section 4.5 shows that this approach leads to sizable power gains in case the innovations are non-Gaussian, while no power is lost under Gaussianity.

4.3.1 A Structural Representation of the Limit Experiment

We consider the limit experiment corresponding to the predictive regression model (4.2.1)–(4.2.2) using the local perturbations (4.2.5)–(4.2.6) and (4.2.8), i.e., the limit of the experiments $\mathcal{E}^{(T)}(f)$ indexed by T , by studying the asymptotic behavior of the induced likelihood ratios. We expand the likelihood ratio around $(\mu, \beta, \gamma, \eta) = (\mu_0, 0, 1, 0)$ and derive its limit in the following proposition.¹⁰

Proposition 4.3.1. *Fix $f \in \mathfrak{F}$. Consider the local parameters $m \in \mathbb{R}$, $b \in \mathbb{R}$, $c \in \mathbb{R}$, and $\eta \in c_{00}$. Then,*

¹⁰As preparation for the results in Section 4.4, we allow in this proposition for local perturbations with respect to γ , even though in the present section γ is assumed to be known.

(i) Under $P_{0,0,0,f}^{(T)}$, the log-likelihood ratio of the predictive regression experiment satisfies, as $T \rightarrow \infty$,

$$\log \frac{dP_{m,b,c,\eta;f}^{(T)}}{dP_{0,0,0,f}^{(T)}} = \Delta^{(T)}(m, b, c, \eta) - \frac{1}{2}Q^{(T)}(m, b, c, \eta) + o_P(1),$$

where

$$\begin{aligned} \Delta^{(T)}(m, b, c, \eta) &= \frac{m}{\sqrt{T}} \sum_{t=2}^T \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) + \frac{b}{T} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) \\ &\quad + \frac{c}{T} \sum_{t=2}^T x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t) + \frac{1}{\sqrt{T}} \sum_{t=2}^T \sum_k \eta_k b_k(y_t^\diamond, \Delta x_t), \\ Q^{(T)}(m, b, c, \eta) &= m^2 J_{f_{yy}} + (2mb J_{f_{yy}} + 2mc J_{f_{yx}}) \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} \\ &\quad + (b^2 J_{f_{yy}} + c^2 J_{f_{xx}} + 2bc J_{f_{yx}}) \frac{1}{T^2} \sum_{t=2}^T \frac{x_{t-1}^2}{\sigma_x^2} \\ &\quad + 2a J'_{f_{yh}} \eta + (2b J'_{f_{yh}} \eta + 2c J'_{f_{xh}} \eta) \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} + \eta' \eta. \end{aligned}$$

(ii) Still under $P_{0,0,0,f}^{(T)}$, as $T \rightarrow \infty$, we have

$$\log \frac{dP_{m,b,c,\eta;f}^{(T)}}{dP_{0,0,0,f}^{(T)}} \Rightarrow \mathcal{L}(m, b, c, \eta) := \Delta(m, b, c, \eta) - \frac{1}{2}Q(m, b, c, \eta), \quad (4.3.1)$$

where

$$\begin{aligned} \Delta(m, b, c, \eta) &= m W_{\ell_{f_y}}(1) + b \int_0^1 W_\varepsilon(s) dW_{\ell_{f_y}}(s) + c \int_0^1 W_\varepsilon(s) dW_{\ell_{f_x}}(s) + \eta' W_h(1), \\ Q(m, b, c, \eta) &= m^2 J_{f_{yy}} + (2mb J_{f_{yy}} + 2mc J_{f_{yx}}) \int_0^1 W_\varepsilon(s) ds \\ &\quad + (b^2 J_{f_{yy}} + c^2 J_{f_{xx}} + 2bc J_{f_{yx}}) \int_0^1 W_\varepsilon(s)^2 ds \\ &\quad + 2m J'_{f_{yh}} \eta + \eta' \eta + (2b J'_{f_{yh}} \eta + 2c J'_{f_{xh}} \eta) \int_0^1 W_\varepsilon(s) ds. \end{aligned}$$

(iii) For every $m, b, c \in \mathbb{R}$ and $\eta \in c_{00}$, $\exp(\mathcal{L}(m, b, c, \eta))$ has unit expectation under $\mathbb{P}_{0,0,0}$.

A proof of Proposition 4.3.1 is provided in Appendix 4.7.2, but let us give a brief sketch here. Part (i) follows from Hallin et al. (2015), which provides generally applicable sufficient conditions for the quadratic expansion of likelihood ratios with densities that are differentiable in quadratic mean (DQM). This DQM condition is implied, for location models, by the absolute continuity of the innovation density function and finiteness of the associated Fisher information, i.e., precisely the content of Assumption 4.2.1. A detailed discussion can be found in Le Cam (1986, Section 17.3) or Le Cam and Yang (2000, Section 7.3). Part (ii) follows from the continuous mapping theorem applied to the weak convergence in (4.2.14). Part (iii) follows from standard stochastic calculations concerning Doléans-Dade exponentials.

Part (iii) of Proposition 4.3.1 ensures that we can introduce a collection of probability measures $\mathbb{P}_{m,b,c,\eta}$ on the measurable space (Ω, \mathcal{F}) (on which the Brownian motions W_ε , $W_{\ell_{fy}}$, $W_{\ell_{fx}}$ and W_h are defined) by the Radon-Nikodym derivative

$$\frac{d\mathbb{P}_{m,b,c,\eta}}{d\mathbb{P}_{0,0,0,0}} = \exp(\mathcal{L}(m, b, c, \eta)), \quad (4.3.2)$$

where $\mathcal{L}(m, b, c, \eta)$ is defined in (4.3.1). Then, in the sense of Hájek-Le Cam (see, for instance, Van der Vaart (2000), Chapter 9), the sequence of predictive regression experiments, indexed by sample size T , weakly converges to the limit experiment described by the measures $\mathbb{P}_{m,b,c,\eta}$. We thus can formally define the limit experiment by

$$\mathcal{E}(f) := \left(\Omega, \mathcal{F}, \left\{ \mathbb{P}_{m,b,c,\eta} : m, b, c \in \mathbb{R}, \eta \in c_{00} \right\} \right), \quad (4.3.3)$$

where $\Omega := C[0, 1] \times C[0, 1] \times C[0, 1] \times C^{\mathbb{N}}[0, 1]$ and $\mathcal{F} := \mathcal{B}_C \otimes \mathcal{B}_C \otimes \mathcal{B}_C \otimes (\otimes_{k=1}^{\infty} \mathcal{B}_C)$.

The following statement is an immediate consequence, by definition, of Proposition 4.3.1.

Corollary 4.3.1. *Let $f \in \mathfrak{F}$, then the sequence of experiments $\mathcal{E}^{(T)}(f)$ converges to the limit experiment $\mathcal{E}(f)$ as $T \rightarrow \infty$.*

Although the log-likelihood ratios $\mathcal{L}(m, b, c, \eta)$ formally describe the limiting experiment, it is more insightful to provide, what we call, a structural representation. This structural representation provides a fixed-horizon continuous time model for which the likelihoods are exactly equal to $\exp(\mathcal{L}(m, b, c, \eta))$. From a statistical point of view, the induced experiments are thus equal. The result follows from an immediate application of Girsanov's theorem to the Radon-Nikodym derivatives (4.3.1). Its proof is therefore omitted.

Theorem 4.3.1. *Fix $f \in \mathfrak{F}$. Let, under $\mathbb{P}_{0,0,0,0}$, Z_ε , $Z_{\ell_{fy}}$, $Z_{\ell_{fx}}$, and Z_h be zero-drift Brownian motions with covariance given by (4.2.15). Then, the limit experiment $\mathcal{E}(f)$ can be described as follows: we observe, on the interval $s \in [0, 1]$, W_ε , $W_{\ell_{fy}}$, $W_{\ell_{fx}}$ and W_h generated by*

$$\begin{aligned} dW_\varepsilon(s) &= cW_\varepsilon(s)ds + dZ_\varepsilon(s), \\ dW_{\ell_{fy}}(s) &= (bJ_{f_{yy}} + cJ_{f_{yx}})W_\varepsilon(s)ds + (mJ_{f_{yy}} + J'_{f_{yh}}\eta)ds + dZ_{\ell_{fy}}(s), \\ dW_{\ell_{fx}}(s) &= (bJ_{f_{yx}} + cJ_{f_{xx}})W_\varepsilon(s)ds + (mJ_{f_{yx}} + J'_{f_{xh}}\eta)ds + dZ_{\ell_{fx}}(s), \\ dW_h(s) &= (bJ_{f_{yh}} + cJ_{f_{xh}})W_\varepsilon(s)ds + (mJ_{f_{yh}} + \eta)ds + dZ_h(s). \end{aligned}$$

A few remarks are to be made concerning Theorem 4.3.1. First, note that for $m = b = c = 0$ and $\eta = 0$, we obtain $W_\varepsilon = Z_\varepsilon$, $W_{\ell_{fy}} = Z_{\ell_{fy}}$, $W_{\ell_{fx}} = Z_{\ell_{fx}}$, and $W_h = Z_h$. Secondly, the theorem essentially states that while $(W_\varepsilon, W_{\ell_{fy}}, W_{\ell_{fx}}, W_h)'$ is an infinite-dimensional zero-drift Brownian motion under $\mathbb{P}_{0,0,0,0}$, it becomes an infinite-dimensional Ornstein-Uhlenbeck process under $\mathbb{P}_{m,b,c,\eta}$, where the log-likelihood ratio $\log(d\mathbb{P}_{m,b,c,\eta}/d\mathbb{P}_{0,0,0,0})$ equals $\mathcal{L}(m, b, c, \eta)$. Finally note that, the second and third stochastic differential equation in Theorem 4.3.1, i.e., those concerning $W_{\ell_{fy}}$ and $W_{\ell_{fx}}$, can be omitted in view of (4.2.16) and (4.2.17). Nevertheless, we keep them here because, on one hand, they describe the limit likelihood ratio when f is known ($\eta = 0$) and, on the other hand, they are useful to describe the likelihood ratio of the maximal invariant \mathcal{M} to be introduced below in (4.3.6).

4.3.2 Maximal Invariant

In the limit experiment $\mathcal{E}(f)$, the parameter $b \in \mathbb{R}$ is the parameter of interest, while the parameters $m \in \mathbb{R}$, $c \in \mathbb{R}$, and $\eta \in c_{00}$ are nuisance parameters. Observe that the nuisance parameters m and η appear only in the drift term of the SDEs in Theorem 4.3.1. This leads directly to an invariance restriction in line with the approach in Zhou, van den Akker, and Werker (2016) for unit root testing.

To be specific, we first introduce, for $m \in \mathbb{R}$ and $\eta \in c_{00}$, the transformations $\mathfrak{g}_{m,\eta} : C^{\mathbb{N}}[0, 1] \rightarrow C^{\mathbb{N}}[0, 1]$ defined by

$$[\mathfrak{g}_{m,\eta}(W)](s) = W(s) - (mJ_{f_y h} + \eta)ds, \quad (4.3.4)$$

for $W \in C^{\mathbb{N}}[0, 1]$ and all $s \in [0, 1]$. Intuitively, the transformation $\mathfrak{g}_{m,\eta}$ adds a drift $s \mapsto -(mJ_{f_y h} + \eta)s$ to W . Thus, Theorem 4.3.1 implies that the law of $(W_\varepsilon, (\mathfrak{g}_{m,\eta}(W_h))')'$ under $\mathbb{P}_{0,b,c,0}$ is the same as the law of $(W_\varepsilon, W_h')'$ under $\mathbb{P}_{m,b,c,\eta}$. By (4.2.16) and (4.2.17), the same holds for $W_{\ell_{fx}}$ and $W_{\ell_{fy}}$. Denote by $\mathfrak{G}_{m,\eta}$ the group of transformations $\mathfrak{g}_{m,\eta}$ for $m \in \mathbb{R}$ and $\eta \in c_{00}$. We can then characterize the maximal invariant with respect to $\mathfrak{G}_{m,\eta}$ in the limit experiment $\mathcal{E}(f)$.

For any process W , we define the associated *bridge process* by

$$B^W(s) := W(s) - sW(1), \quad (4.3.5)$$

for all $s \in [0, 1]$. Then, one readily verifies

$$\begin{aligned} B^{\mathfrak{g}_{m,\eta}(W)}(s) &= [\mathfrak{g}_{m,\eta}(W)](s) - s[\mathfrak{g}_{m,\eta}(W)](1) \\ &= W(s) - (mJ_{f_y h} + \eta)ds - s(W(1) - (mJ_{f_y h} + \eta)) \\ &= W(s) - sW(1) \\ &= B^W(s). \end{aligned}$$

As a result, the bridges $B^{W_{\ell_{fy}}}$, $B^{W_{\ell_{fx}}}$, and B^{W_h} , as well as W_ε , are invariant under the transformations $\mathfrak{g}_{m,\eta}$.

Define the mapping M by $M(W_\varepsilon, W_h) := (W_\varepsilon, B_h)$ with $B_h = B^{W_h}$. It then follows that statistics that are measurable with respect to the σ -field

$$\mathcal{M} = \sigma(M(W_\varepsilon, W_h)) = \sigma(W_\varepsilon, B_h) \quad (4.3.6)$$

are invariant with respect to $\mathfrak{g}_{m,\eta}$ for all $m \in \mathbb{R}$ and $\eta \in c_{00}$. Moreover, in the following theorem, we show \mathcal{M} to be maximally invariant. Its proof is again provided in the appendix.

Theorem 4.3.2. *In the limit experiment $\mathcal{E}(f)$, for $m \in \mathbb{R}$ and $\eta \in c_{00}$, the σ -field \mathcal{M} in (4.3.6) is the maximal invariant with respect to $\mathfrak{G}_{m,\eta}$.*

4.3.3 Semiparametric Power Envelope

Theorem 4.3.2 implies that any inference invariant with respect to $\mathfrak{G}_{m,\eta}$ must be a measurable with respect to \mathcal{M} (see, e.g., Lehmann and Romano (2005, Theorem 6.2.1)). Therefore, by the Neyman-Pearson lemma, inference based on likelihood ratio with respect to \mathcal{M} yields the power envelope for the invariant tests in the limit experiment $\mathcal{E}(f)$. The following result provides this likelihood ratio.

Theorem 4.3.3. *Fix $f \in \mathfrak{F}$. Then the likelihood ratios in the limit experiment $\mathcal{E}(f)$ restricted to the maximal invariant \mathcal{M} are given by*

$$\frac{d\mathbb{P}_{b,c}^{\mathcal{M}}}{d\mathbb{P}_{0,0}^{\mathcal{M}}} := \mathbb{E} \left[\frac{d\mathbb{P}_{m,b,c,\eta}}{d\mathbb{P}_{0,0,0,0}} | \mathcal{M} \right] = \exp \left(\Delta_{\mathcal{M}}(b, c) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}}(b, c) \right), \quad (4.3.7)$$

where

$$\begin{aligned} \Delta_{\mathcal{M}}(b, c) &= b \int_0^1 W_{\varepsilon}(s) dB_{\ell_{fy}}(s) + c \int_0^1 W_{\varepsilon}(s) dB_{\ell_{fx}}(s) + c W_{\varepsilon}(1) \overline{W}_{\varepsilon}, \\ \mathcal{Q}_{\mathcal{M}}(b, c) &= (b^2 J_{f_{yy}} + c^2 J_{f_{xx}} + 2bc J_{f_{yx}}) \left[\overline{W}_{\varepsilon}^2 - (\overline{W}_{\varepsilon})^2 \right] + c^2 (\overline{W}_{\varepsilon})^2. \end{aligned}$$

with $\overline{W}_{\varepsilon} = \int_0^1 W_{\varepsilon}(s) ds$ and $\overline{W}_{\varepsilon}^2 = \int_0^1 W_{\varepsilon}^2(s) ds$.

The restriction to invariant tests, removes the nuisance parameters m and η from the testing problem. Indeed, the likelihood ratio (4.3.7) no

longer depends on either m or η . We formally define the limit experiment restricted to the maximal invariance \mathcal{M} as follows.

$$\mathcal{E}_{\mathcal{M}}(f) := \left(\Omega, \mathcal{M}, \left\{ \mathbb{P}_{b,c} : b, c \in \mathbb{R} \right\} \right). \quad (4.3.8)$$

Again, the likelihood ratios $d\mathbb{P}_{b,c}^{\mathcal{M}}/d\mathbb{P}_{0,0}^{\mathcal{M}}$ can also be interpreted as Girsanov transformation. We state this as a corollary as the result follows immediately from calculating the bridges corresponding to $W_{\ell_{fy}}$ and $W_{\ell_{fx}}$ in Theorem 4.3.1.

Corollary 4.3.2. *Fix $f \in \mathfrak{F}$. Let, under $\mathbb{P}_{b=0,c=0}^{\mathcal{M}}$, Z_{ε} , $Z_{\ell_{fy}}$, and $Z_{\ell_{fx}}$ be zero-drift Brownian motions with covariance given by the upper-left three-by-three block of (4.2.15). Then, the limit experiment $\mathcal{E}_{\mathcal{M}}(f)$ can be described as follows: we observe, on the interval $s \in [0, 1]$, W_{ε} , $B_{\ell_{fy}}$, and $B_{\ell_{fx}}$ defined by*

$$\begin{aligned} dW_{\varepsilon}(s) &= cW_{\varepsilon}(s)ds + dZ_{\varepsilon}(s), \\ dB_{\ell_{fy}}(s) &= (bJ_{f_{yy}} + cJ_{f_{yx}})W_{\varepsilon}^{\mu}(s)ds + d[Z_{\ell_{fy}}(s) - sZ_{\ell_{fy}}(1)], \\ dB_{\ell_{fx}}(s) &= (bJ_{f_{yx}} + cJ_{f_{xx}})W_{\varepsilon}^{\mu}(s)ds + d[Z_{\ell_{fx}}(s) - sZ_{\ell_{fx}}(1)], \end{aligned}$$

where $W_{\varepsilon}^{\mu}(s) = W_{\varepsilon}(s) - \overline{W_{\varepsilon}}$.

Corollary 4.3.2 does not provide, as far as we know, a further invariance structure that can be used to eliminate the nuisance parameter c . As a result, we rely, in Section 4.4, on the so-called Approximate Least Favorable Distribution method to deal with this final nuisance parameter.

We conclude this section by considering the special case where the innovation density f is Gaussian.

Remark 4.3.1 (Gaussian case). When the innovation density f is Gaussian, the maximal invariant \mathcal{M} degenerates to σ -field $\mathcal{M}_{\varepsilon} = \sigma(W_{\varepsilon})$. The likelihood ratio based on $\mathcal{M}_{\varepsilon}$, of course, coincides with the Gaussian limit likelihood ratio in Lemma 3 of Jansson and Moreira (2006).

We can now consider the the log-likelihood ratio tests, for given $c \in (-\infty, 0]$,

$$\varphi(b, c) = \mathbb{1}\{\Delta_{\mathcal{M}}(b, c) - \frac{1}{2}\mathcal{Q}_{\mathcal{M}}(b, c) > \kappa(b, c, f; \alpha)\}, \quad (4.3.9)$$

where $\kappa(b, c, f; \alpha)$ is implicitly defined by the level restriction $\mathbb{E}[\varphi(b, c)] = \alpha$. Then, by the Neyman-Pearson lemma, we know that the test $\varphi(\bar{b}, c)$ is point-optimal for the hypothesis $H_0 : b = 0$ versus $H_1 : b = \bar{b}$, under the assumption of a given c . The power envelope for invariant tests in $\mathcal{E}(f)$ is then given by

$$\Psi(b, c) := \mathbb{E} \left[\varphi(b, c) \frac{d\mathbb{P}_{m,b,c,\eta}}{d\mathbb{P}_{0,0,0,0}} \right]. \quad (4.3.10)$$

Furthermore, by the Asymptotic Representation Theorem (see Van der Vaart (2000, Chapter 9)), $\Psi(b, c)$ is the asymptotic power envelope for all asymptotically invariant tests in $\mathcal{E}^{(T)}(f)$.

4.3.4 Rank-based asymptotically invariant statistics

The elimination of the nuisance parameters m and η is performed in the limit experiment $\mathcal{E}(f)$ and leads to $\mathcal{E}_{\mathcal{M}}(f)$. We now show how this elimination can be mimicked in the actual predictive regression model of interest. It is reasonable to expect that exploiting the asymptotic invariance structures, also works “well” for the sequence of experiments $\mathcal{E}^{(T)}(f)$. The quality of this approximation will be assessed by simulation in Section 4.5.

The appearance of the Brownian Bridges $B_{\ell_{f_x}}$ and $B_{\ell_{f_y}}$ in Corollary 4.3.2, naturally suggest to use statistics that are based on ranks of the innovations ε_t^x and ε_t^y in the predictive regression model. Indeed, we will follow that route now. A complication, with respect to the traditional rank-based literature is that the innovations are bivariate.

Let, for $i \in \{y, x\}$, g_i be so-called (marginal) *reference densities*. These (marginal) reference densities can be freely chosen by the researcher, subject to the following assumption.

Assumption 4.3.1. The reference densities g_i , $i = \{y, x\}$, are strictly positive, absolutely continuous with derivative \dot{g}_i and $J_{g_i} := \int (\dot{g}_i/g_i)^2 g_i < \infty$. Moreover, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left(-\frac{\dot{g}_i}{g_i} \left[G_i^{-1} \left(\frac{t}{T+1} \right) \right] \right)^2 = J_{g_i}, \quad (4.3.11)$$

where G_i^{-1} is the inverse cumulative distribution function associated to $g_i(\cdot)$.

Now let $R_{x,t}$ denote the ranks of $x_t - x_{t-1}$, while $R_{y,t}$ denotes the rank of y_t . Note that the pairs $(R_{y,t}, R_{x,t})$ equal the (component-wise) ranks of $(\varepsilon_t^y, \varepsilon_t^x)$ under $\beta = 0$ and $\gamma = 1$. Then, we introduce the reference score function

$$\begin{aligned} \ell_g(u_1, u_2) &= (\ell_{g_y}(u_1), \ell_{g_x}(u_2))' \\ &:= \mathbf{R}^{-1} \left(-\frac{\dot{g}_y}{g_y} [G_y^{-1}(u_1)], -\frac{\dot{g}_x}{g_x} [G_x^{-1}(u_2)] \right)', \end{aligned} \quad (4.3.12)$$

where $\mathbf{R} := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ is a chosen reference correlation matrix. The construction of the score ℓ_g is motivated by the score function in the Gaussian case, i.e., $\ell_\phi(\varepsilon_t) = \mathbf{R}^{-1} (\varepsilon_t^y/\sigma_y, \varepsilon_t^x/\sigma_x)'$. Recall that for Gaussian densities g , we have that $-\dot{g}g$ equals the identity. Thus, we consider the (component-wise) rank-based scores $-\frac{\dot{g}_i}{g_i} [G_i^{-1}(\frac{R_{i,t}}{T+1})]$. This structure also resembles the idea of using a Gaussian copula, where the dependence is captured by the correlation matrix \mathbf{R} .

Finally, for $s \in [0, 1]$, define the partial sum process of the rank-based scores by

$$B_{\ell_g}^{(T)}(s) = \left(B_{\ell_{g_y}}^{(T)}(s), B_{\ell_{g_x}}^{(T)}(s) \right)' := \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \ell_g \left(\frac{R_{y,t}}{T+1}, \frac{R_{x,t}}{T+1} \right). \quad (4.3.13)$$

The following result establishes the limiting behavior of $B_{\ell_g}^{(T)}$ under $\mathbf{P}_{m,0,0,\eta;f}^{(T)}$. Its proof is again provided in Appendix 4.7.2.

Proposition 4.3.2. *Suppose $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)'$ are i.i.d innovations with density $f \in \mathfrak{F}$. Let g_y and g_x be reference densities that satisfy Assumption 4.3.1. Then, under $P_{m,0,0,\eta;f}^{(T)}$, we have*

$$B_{\ell_g}^{(T)} \Rightarrow B_{\ell_g}, \quad (4.3.14)$$

where B_{ℓ_g} is a bivariate Brownian bridge, i.e., $B_{\ell_g}(s) = W_{\ell_g}(s) - sW_{\ell_g}(1)$, with W_{ℓ_g} a Brownian motion. The covariance of W_{ℓ_g} with W_ε and $W_{\ell_f} := (W_{\ell_{fy}}, W_{\ell_{fx}})'$ is given by

$$\text{Var} \begin{pmatrix} W_\varepsilon(1) \\ W_{\ell_f}(1) \\ W_{\ell_g}(1) \end{pmatrix} = \begin{pmatrix} 1 & e'_1 & \sigma'_{\varepsilon g} \\ e_1 & J_f & J_{fg} \\ \sigma_{\varepsilon g} & J_{gf} & J_g \end{pmatrix}, \quad (4.3.15)$$

where

$$\begin{aligned} e_1 &= (0, 1)', \\ \sigma_{\varepsilon g} &= (\sigma_{\varepsilon g_y}, \sigma_{\varepsilon g_x})' = E_f \varepsilon_t^x \ell_g(F_y(\varepsilon_t^y), F_x(\varepsilon_t^x)), \\ J_{fg} &= J'_{gf} = E_f \ell_f(F_y(\varepsilon_t^y), F_x(\varepsilon_t^x)) \ell_g(F_y(\varepsilon_t^y), F_x(\varepsilon_t^x))', \\ J_g &= E_f \ell_g(F_y(\varepsilon_t^y), F_x(\varepsilon_t^x)) \ell_g(F_y(\varepsilon_t^y), F_x(\varepsilon_t^x))'. \end{aligned}$$

The above result is classical for univariate rank statistics. In the present paper, we use component-wise bivariate ranks. A similar idea is used in Theorem 3.1 of Zhou (2017).

We will use the rank-based processes $B_{\ell_g}^{(T)}$ to replace B_{ℓ_f} in likelihood ratio in Theorem 4.3.3. One could contemplate to use reference densities \hat{f} based on a non-parametric estimate of the true innovation density, but we leave that for future work. As we will see in Section 4.5, even for incorrectly chosen reference densities (that is, for $g \neq f$), our procedure features power gains over existing Gaussian based procedures. These gains come from the maintained assumption that the innovations ε_t are i.i.d. Such an assumption is often maintained in empirical work. It's important to note that choosing a reference density $g \neq f$ does not affect the validity of our test. The test will be of the appropriate size irrespective of g .

The behavior of B_{ℓ_g} under the measure $\mathbb{P}_{m,b,c,\eta}$ are provided in the corollary below. Based on this result, we can capture the limit behavior of statistic $B_{\ell_g}^{(T)}$ under the measure $\mathbb{P}_{m,0,c,\eta}$, which is necessary for the ALFD approach introduced in the section below to eliminate nuisance parameter b and to simulate the critical value.

Corollary 4.3.3. *Fix $f \in \mathfrak{F}$. Let $m, b \in \mathbb{R}$, $c \in (-\infty, 0]$, and $\eta \in c_{00}$. Then, under $\mathbb{P}_{m,b,c,\eta}$, the behavior of B_{ℓ_g} follows*

$$dB_{\ell_g}(s) = J_{gf} \begin{pmatrix} b \\ c \end{pmatrix} W_{\varepsilon}^{\mu}(s) ds + d[Z_{\ell_g}(s) - sZ_{\ell_g}(1)], \quad (4.3.16)$$

where Z_{ℓ_g} is a bivariate Brownian motion with variance J_g and covariance with Z_{ε} equal to $\sigma_{\varepsilon g}$.

Remark 4.3.2. In addition, Corollary 4.3.3 also intuitively reveals how the rank-based quasi-likelihood ratio (QLR) test constructed with $B_{\ell_g}^{(T)}$ has more power than the canonical Gaussian QLR test. The latter is constructed based on the assumption that f is Gaussian, with which, we have $J_f = \mathbf{R}^{-1}$ and $W_{\ell_f} = \mathbf{R}^{-1}(W_{\varepsilon}, W_{\varepsilon^y})'$, where W_{ε^y} is the limit of partial-sum process $T^{-1/2} \sum_{t=2}^{\lfloor sT \rfloor} y_t^{\diamond} / \sigma_y$. Then, the results in Corollary 4.3.2 becomes

$$\mathbf{R}^{-1} d \begin{pmatrix} B_{\varepsilon}(s) \\ B_{\varepsilon^y}(s) \end{pmatrix} = \mathbf{R}^{-1} \begin{pmatrix} b \\ c \end{pmatrix} W_{\varepsilon}^{\mu}(s) ds + \mathbf{R}^{-1} d \begin{pmatrix} Z_{\varepsilon}(s) - sZ_{\varepsilon}(1) \\ Z_{\varepsilon^y}(s) - sZ_{\varepsilon^y}(1) \end{pmatrix},$$

where $B_{\varepsilon^y}(s) := W_{\varepsilon^y}(s) - sW_{\varepsilon^y}(1)$.¹¹

Since the source of power gain comes from the shift term of these SDEs, the magnitudes of these shifts mainly determine the performances of different tests. Thus, when f is not Gaussian, with properly chosen reference marginal densities g_y and g_x , often we have $J_{gf} \geq \mathbf{R}^{-1}$. This induces a higher power of the rank-based QLR test than the Gaussian QLR test.

¹¹Note that this is also a immediate result of $dW_{\varepsilon}(s) = cW_{\varepsilon}(s)ds + dZ_{\varepsilon}(s)$ by taking bridge and premultiplying \mathbf{R}^{-1} , indicating that the Gaussian quasi-likelihood test only uses information contained in $\sigma(W_{\varepsilon})$.

Moreover, if we fix both g_y and g_x to be Gaussian, then we have for any true density f that $J_{fg}\mathbf{R} = \text{diag}\{J_{f_y g_y}, J_{f_x g_x}\}$ with $J_{f_y g_y} \geq 1$ and $J_{f_x g_x} \geq 1$ (thus $J_{gf} \geq \mathbf{R}^{-1}$). These equalities hold if f is also Gaussian. This result induces the performance of our rank-based inference in Figure 4.2-4.3 in Section 4.5 which is related to the traditional Chernoff-Savage result.

In any case, likelihood ratio tests based on Theorem 4.3.3 still feature the nuisance parameter c . We deal with this in the next section.

4.4 Eliminating the nuisance parameter γ by ALFD

In the previous section, we have developed the semiparametric power envelope for tests that are asymptotically invariant with respect to m and η , under the assumption that γ is known. We now address the question in case γ is treated as a nuisance parameter.

As argued in the the discussion following Corollary 4.3.2, we conjecture that the nuisance parameter c cannot be dealt with using invariance arguments. Various alternative methods to deal with nuisance parameters in testing problems have been used in the literature. In relation to the predictive regression model at hand here, we mention the Bonferroni method (Cavanagh, Elliott, and Stock (1995) and Campbell and Yogo (2005)); tests based on a conditional unbiasedness condition (Jansson and Moreira (2006)); and tests based on a numerically calculated Approximate Least Favorable Distribution (ALFD) as more recently proposed in Elliott, Müller, and Watson (2015).

These approaches have different advantages and disadvantages. Inference based on Bonferroni bounds is simple, but can be severely undersized when the predictor is “far away” from being a unit root ($\gamma \ll 1$). In such a case, confidence intervals obtained by inverting the test may end up having essentially zero coverage probability (see Phillips (2014)). The Jansson and Moreira (2006) test is optimal in the class of conditional unbiased tests,

however, its simulation results show that it has relatively low power compared to the Campbell and Yogo (2005) test based on a modified Bonferroni method.

We will follow the ALFD approach introduced by Elliott, Müller, and Watson (2015). This leads to tests that are of correct size for all relevant c and have good power performance compared with the other approaches. This will be confirmed by simulations in Section 4.5, where we show that these properties still hold after we have reduced the testing problem by invariance arguments as in Section 4.3. That is, our test enjoys the same size properties but with improved power for non-Gaussian innovations.

4.4.1 The Approximately Least Favorable Distribution (ALFD) Approach

In order to apply the ALFD approach to the experiment $\mathcal{E}_{\mathcal{M}}(f)$, observe that we can rewrite its associated log-likelihood ratios in Theorem 4.3.3 as

$$\mathcal{L}_{\mathcal{M}}(b, c) := bS_1 + cS_2 - \frac{1}{2} [(b, c)J_f(b, c)' - c^2] S_3 - \frac{1}{2}c^2S_4 \quad (4.4.1)$$

where

$$\begin{aligned} S_1 &= \int_0^1 W_{\varepsilon}(s) dB_{\ell_{fy}}(s), \quad S_2 = \int_0^1 W_{\varepsilon}(s) dB_{\ell_{fx}}(s) + W_{\varepsilon}(1)\overline{W}_{\varepsilon}, \\ S_3 &= \overline{W_{\varepsilon}^2} - (\overline{W_{\varepsilon}})^2 \quad \text{and} \quad S_4 = \overline{W_{\varepsilon}^2}. \end{aligned} \quad (4.4.2)$$

We can thus consider the four-dimensional sufficient statistic $S = (S_1, S_2, S_3, S_4)$, whose joint distribution under $\mathbb{P}_{m,b,c,\eta}$ we denote by $F_{b,c}(S)$. Observe that, in view of Corollary 4.3.2, this distribution indeed only depends on the parameter of interest b and the nuisance parameter c , but not on m or η .

The hypothesis of interest is

$$H_0 : b = 0, \quad c \in (-\infty, 0] \quad \text{versus} \quad H_1 : b > 0, \quad c \in (-\infty, 0]. \quad (4.4.3)$$

Note that, thus, both the null and the alternative hypothesis are composite. We first discuss elimination of the nuisance parameter c under the alternative and, subsequently, its elimination under the null.

To eliminate the nuisance parameter c under the alternative, a standard approach is to consider a so-called weighted average power (see, e.g., Andrews and Ploberger (1994))

$$\text{WAP}(\varphi) = \int \left(\int \varphi(S) dF_{b,c}(S) \right) d\Lambda_1(c), \quad (4.4.4)$$

where φ is some test function for the problem above and Λ_1 is a probability weighting measure related to c with support on its parameter space. The weighting measure Λ_1 can be chosen by the researcher and reflects the weights that she assigns to various values of c under the alternative. Due to Fubini's Theorem, we have

$$\text{WAP}(\varphi) = \int_c \int_S \varphi(S) dF_{b,c}(S) d\Lambda_1(c) = \int_S \varphi(S) d \int_c F_{b,c}(S) d\Lambda_1(c), \quad (4.4.5)$$

which leads to the simple alternative hypothesis $H_{1;\Lambda_1}$, under which the distribution of S is given by $F_{b;\Lambda_1}(S) = \int F_{b,c}(S) d\Lambda_1(c)$. In this way, the testing problem is reduced to testing H_0 against $H_{1;\Lambda_1}$.

Subsequently, in order to eliminate the nuisance parameter c under the null we proceed as follows. Again we impose a probability weighting measure Λ_0 for c and introduce the simple null hypothesis, denoted $H_{0;\Lambda_0}$, under which the distribution of S is given by $F_{b;\Lambda_0}(S) = \int F_{b,c}(S) d\Lambda_0(c)$. Now we define the test $\varphi_{\bar{b};\Lambda}$ by

$$\varphi_{\bar{b},\Lambda_0}(S) = \begin{cases} 1 & \text{if } dF_{\bar{b},\Lambda_1}(S) > \kappa dF_{0,\Lambda_0}(S), \\ 0 & \text{if } dF_{\bar{b},\Lambda_1}(S) < \kappa dF_{0,\Lambda_0}(S), \end{cases} \quad (4.4.6)$$

where the critical value κ is chosen to obtain the desired size. By the Neyman-Pearson Lemma, $\varphi_{\bar{b},\Lambda_0}$ is the point-optimal, at $b = \bar{b}$, for the problem of testing the null $H_{0;\Lambda_0}$ against the alternative $H_{1;\Lambda_1}$.

The problem of choosing Λ_0 is, unfortunately, more complicated than that of choosing Λ_1 . The reason is that we want to control the rejection probability of the test, not only under $H_{0;\Lambda_0}$, but for all values of $c \in (-\infty, 0]$. In general there is no reason to expect that a level- α test under $H_{0;\Lambda_0}$ is of correct size for the entire null hypothesis H_0 . However, for some specific choices of Λ_0 this statement is true, and such a distribution is called a *least favorable*

distribution; see, e.g., Lehmann and Romano (2005), Theorem 3.8.1. Formally, a distribution Λ_0^* is called *least favorable* if the most powerful level- α test for testing $H_{0;\Lambda_0^*}$ against $H_{1;\Lambda_1}$ is of size $\leq \alpha$ for the (entire) null hypothesis H_0 . Moreover, once more by Theorem 3.8.1 in Lehmann and Romano (2005), the test $\varphi_{\bar{b},\Lambda_0^*}$ is also point optimal (at $b = \bar{b}$) for this problem. A least favorable distribution Λ_0^* exists in most of the usual statistical problems. conditions that ensure this and associated references can be found in Section 3.8 of Lehmann and Romano (2005).

As, in most cases, the least favorable distribution Λ_0^* is not easily obtained, Elliott, Müller, and Watson (2015) propose a numerical method to find, what they call, an “Approximate Least Favorable Distribution” (ALFD). The ALFD is defined as follows.

Definition 4.4.1. An ϵ -ALFD is a probability distribution $\Lambda_0^{*\epsilon}$ over $(-\infty, 0]$ satisfying

- (i) the Neyman-Pearson test (4.4.6) with $\Lambda = \Lambda_0^{*\epsilon}$ and critical value $\kappa = \kappa^*$, i.e., $\varphi_{\bar{b},\Lambda_0^{*\epsilon}}$, is of size α under $H_{0;\Lambda_0^{*\epsilon}}$ and has power $\bar{\pi}$ against $H_{1;\Lambda_1}$;
- (ii) there exists $\kappa^{*\epsilon}$ such that the test (4.4.6) with $\Lambda = \Lambda_0^{*\epsilon}$ and $\kappa = \kappa^{*\epsilon}$, $\varphi_{\bar{b},\Lambda_0^{*\epsilon}}^\epsilon$, is of level α under H_0 , and has power of at least $\bar{\pi} - \epsilon$ against $H_{1;\Lambda_1}$.

The test $\varphi_{\bar{b},\Lambda_0^{*\epsilon}}^\epsilon$ (in particular, the ALFD $\Lambda_0^{*\epsilon}$ and the critical value $\kappa^{*\epsilon}$) is exactly we are looking for, once we have set the weights Λ_1 of interest for the alternative hypothesis. Besides the size control under H_0 , the definition above also ensures that the test $\varphi_{\bar{b},\Lambda_0^{*\epsilon}}^\epsilon$ enjoys a nearly-optimal property with a relatively small power loss (less than ϵ).

Note that even for a given (small) value of ϵ , the ALFD $\Lambda_0^{*\epsilon}$ is not necessarily “close” to the least favorable distribution Λ_0^* . Actually, (possibly infinitely) many pairs of $(\Lambda_0^{*\epsilon}, \kappa^{*\epsilon})$ will satisfy Definition 4.4.1. The details about how to implement the numerical algorithm to determine a pair of

$(\Lambda_0^{*\epsilon}, \kappa^{*\epsilon})$ (henceforth the test $\varphi_{\bar{b}, \Lambda_0^{*\epsilon}}$) for a small ϵ can be found in Section 3 and Appendix A of Elliott, Müller, and Watson (2015).

4.4.2 Putting it all together

Asymptotically invariant tests, with respect to m and η , can be based on Theorem 4.3.3 using the log-likelihood ratio (4.4.1). As the innovation distribution is unknown, we use Proposition 4.3.2 to replace $B_{\ell_{fy}}$ and $B_{\ell_{fx}}$ by $B_{\ell_{gy}}^{(T)}$ and $B_{\ell_{gx}}^{(T)}$, respectively. More precisely, we consider the rank-based finite-sample counterpart of (4.4.1), i.e., we replace (4.4.2) by

$$\widehat{S}^g := (\widehat{S}_1^g, \widehat{S}_2^g, \widehat{S}_3^g, \widehat{S}_4^g), \quad (4.4.7)$$

where

$$\begin{aligned} \widehat{S}_1^g &= \int_0^1 W_\varepsilon^{(T)}(s) dB_{\ell_{gy}}^{(T)}(s), \\ \widehat{S}_2^g &= \int_0^1 W_\varepsilon^{(T)}(s) dB_{\ell_{gx}}^{(T)}(s) + W_\varepsilon^{(T)}(1) \int_0^1 W_\varepsilon^{(T)}(s) ds, \\ \widehat{S}_3^g &= \int_0^1 \left(W_\varepsilon^{(T)}(s) \right)^2 ds - \left(\int_0^1 W_\varepsilon^{(T)}(s) ds \right)^2, \\ \widehat{S}_4^g &= \int_0^1 \left(W_\varepsilon^{(T)}(s) \right)^2 ds. \end{aligned}$$

Using the continuous mapping theorem and Proposition 4.3.2, we have, under $P_{m,0,0,\eta;f}^{(T)}$,

$$\widehat{S}^g \Rightarrow S^g := (S_1^g, S_2^g, S_3^g, S_4^g), \quad (4.4.8)$$

where

$$\begin{aligned} S_1^g &= \int_0^1 W_\varepsilon(s) dB_{\ell_{gy}}(s), \quad S_2^g = \int_0^1 W_\varepsilon(s) dB_{\ell_{gx}}(s) + W_\varepsilon(1) \overline{W_\varepsilon}, \\ S_3^g &= \overline{W_\varepsilon^2} - (\overline{W_\varepsilon})^2 \quad \text{and} \quad S_4^g = \overline{W_\varepsilon^2}. \end{aligned} \quad (4.4.9)$$

The behavior of S^g under $\mathbb{P}_{m,b,c,\eta}$ follows from Corollary 4.3.3.

Based on the rank-based statistics (in the limit) $S^g = (S_1^g, S_2^g, S_3^g, S_4^g)$ defined in (4.4.8) and (4.4.9), we introduce our rank-based quasi-likelihood statistic

$$\mathcal{L}_g(b, c) = bS_1^g + cS_2^g - \frac{1}{2} [(b, c)J_p(b, c)' - c^2] S_3^g - \frac{1}{2} c^2 S_4^g \quad (4.4.10)$$

with a chosen Fisher information

$$J_p = \begin{pmatrix} J_{p_{yy}} & J_{p_{yx}} \\ J_{p_{yx}} & J_{p_{xx}} \end{pmatrix} := \mathbf{R}^{-\frac{1}{2}} \text{diag}\{J_{g_y}, J_{g_x}\} \mathbf{R}^{-\frac{1}{2}'} \quad (4.4.11)$$

This choice of the Fisher information J_p is based on the fact that $\mathbf{R}^{\frac{1}{2}} J_f \mathbf{R}^{\frac{1}{2}'}$ is a diagonal matrix with elements larger or equal to 1 (see Kagana and Landsmanb (1999)).

Now, applying the ALFD algorithm to \mathcal{L}_g , we obtain a distribution $\Lambda_{0,g}^{*\epsilon}$ and critical value $\kappa_{g,n}$ such that the test

$$\varphi_{g,n}(S^g, \rho) = \begin{cases} 1 & \text{if } \int \mathcal{L}_g(\bar{b}, c) d\Lambda_1(c) > \kappa_{g,n} \int \mathcal{L}_g(0, c) d\Lambda_{0,g}^{*\epsilon}(c) \\ 0 & \text{if } \int \mathcal{L}_g(\bar{b}, c) d\Lambda_1(c) < \kappa_{g,n} \int \mathcal{L}_g(0, c) d\Lambda_{0,g}^{*\epsilon}(c) \end{cases} \quad (4.4.12)$$

is of size α . Here \bar{b} serves as a fixed alternative point for the quasi-likelihood statistic (c.f. Elliott, Rothenberg, and Stock (1996)).

In order to make the above test feasible, we make the following assumption.

Assumption 4.4.1. For all $f \in \mathfrak{F}$,

- (a) There exists some estimators $\hat{\sigma}_y$, $\hat{\sigma}_x$ and $\hat{\rho}$ of parameters σ_y , σ_x and ρ respectively, satisfying that, under $P_{m,0,c,\eta;f}^{(T)}$, $\hat{\sigma}_y \xrightarrow{p} \sigma_y$, $\hat{\sigma}_x \xrightarrow{p} \sigma_x$ and $\hat{\rho} \xrightarrow{p} \rho$, as $T \rightarrow \infty$.
- (b) There exists some estimators $\hat{\sigma}_{\varepsilon g_x}$, \hat{J}_g and \hat{J}_{fg} of parameters $\sigma_{\varepsilon g_x}$, J_g and J_{fg} respectively, satisfying that, under $P_{m,0,c,\eta;f}^{(T)}$, $\hat{\sigma}_{\varepsilon g_x} \xrightarrow{p} \sigma_{\varepsilon g_x}$, $\hat{J}_g \xrightarrow{p} J_g$ and $\hat{J}_{fg} \xrightarrow{p} J_{fg}$, as $T \rightarrow \infty$.

The existence of estimators mentioned in Part (a) of Assumption 4.4.1 is standard. In the Monte-Carlo study of Section 4.5, candidates for $\hat{\sigma}_y$, $\hat{\sigma}_x$ and $\hat{\rho}$ are provided. The existence of estimators in Part (b) ensures the feasibility of the numerically determined pair $(\Lambda_0^{*\epsilon}, \kappa^{*\epsilon})$. In applications these can easily

be estimated, however in the Monte Carlo study we assume $\sigma_{\varepsilon g_x}$, J_g and J_{fg} to be known. This is necessary as we cannot afford to determine a pair $(\Lambda_0^{*\epsilon}, \kappa^{*\epsilon})$ for each repetition in the simulation. That would be too intensive computationally.

4.5 A Monte Carlo Study

In this section, we explore by Monte Carlo the size and power properties of our test (4.4.12), combined with the switching approach detailed in Appendix 4.8, (labeled “WZ”) relative to the Gaussian quasi-likelihood counterpart in Elliott, Müller, and Watson (2015) (labeled “EMW”). From the theoretical results, both tests should enjoy good size properties but the WZ-test should exhibit larger power in case the true innovation distribution F is not Gaussian. Under Gaussianity, both tests should have similar power. In our simulation setup, we follow Jansson and Moreira (2006). More precisely, we simulate the model (4.2.1)–(4.2.2) with $\mu = 2$, $\sigma_y = 4$, $\sigma_x = 3$, and $\rho = -0.5$. All results reported in this section are based on 10,000 replications.

Following Elliott, Müller, and Watson (2015), for the ALFD approach: we choose a discrete weighting distribution Λ_1 in (4.4.4) where each of the 51 points

$$c \in \{0, -0.25^2, -0.5^2, \dots, -12.5^2\}$$

of the support have equal weight. The same 51 points are also as the support of $\Lambda_0^{*\epsilon}$. In order to have roughly similar power for each value of c , again following Elliott, Müller, and Watson (2015), we transform the parameter δ by

$$b = R(\delta) = \delta \sqrt{\frac{-2c + 6}{1 - \rho^2}}, \quad \text{for } c < 0. \quad (4.5.1)$$

Alternatives for β are now characterized by different values of δ . Finally, for the test statistic in (4.4.12), we choose a fixed alternative $\bar{b} = R(1.645)$ where the power is about 50%.

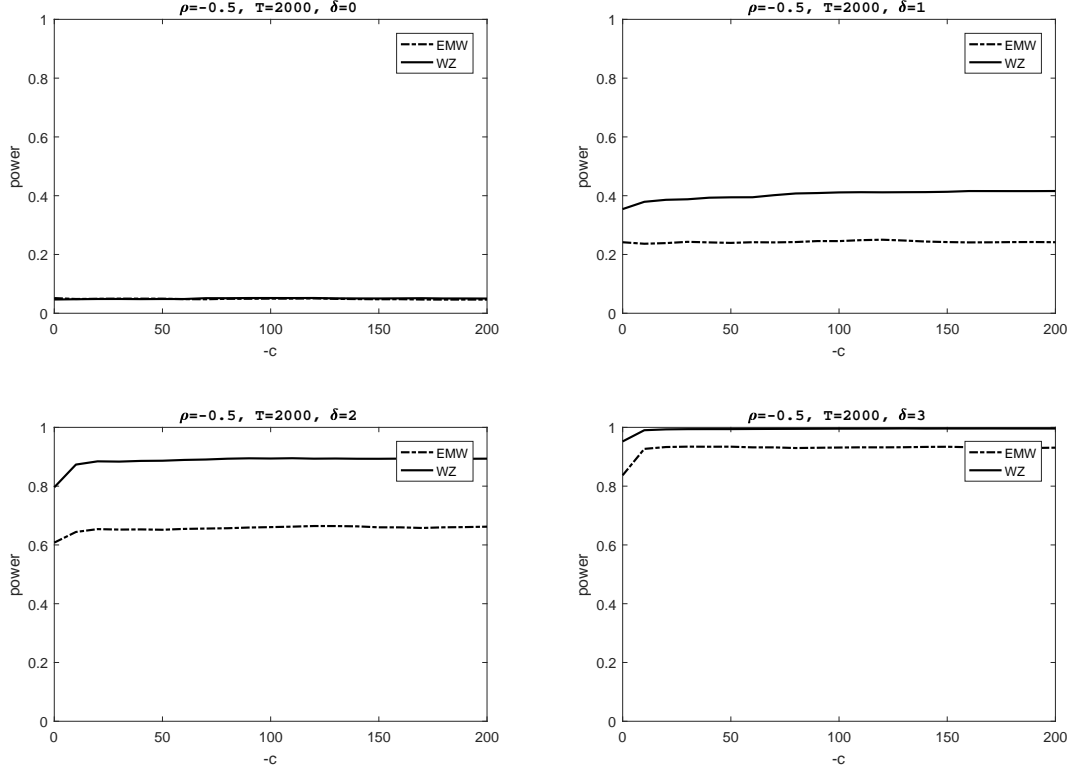


Figure 4.1: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate t_3 and Power functions of the EMW test, under the true innovation densities f is Multivariate t_3 . The sample size T is 2000.

To plot the power functions, we let the local parameter c (which governs the persistent level of the regressor) take 21 values $c \in \{0, -10, -20, \dots, -200\}$ and we let the parameter of interest b take four values $b = B(\delta)$ with $\delta \in \{0, 1, 2, 3\}$. The null hypothesis H_0 corresponds to $\delta = 0$. The significance level α is chosen to be 5%.

In Figures 4.1–4.4, we reports the large-sample ($T = 2000$) size and power properties of our rank-based WZ test and the EMW test, for different combinations of the true density f and the marginal reference densities g_y and g_x . Figure 4.1 reports the case where f is a Multivariate t_3 density, while g_y and g_x are both univariate t_3 densities. Both the WZ test and the EMW test are of correct size for all chosen values of c . Under the alternative hypothesis (i.e., for $\delta \in \{1, 2, 3\}$), the see that the WZ test is more powerful

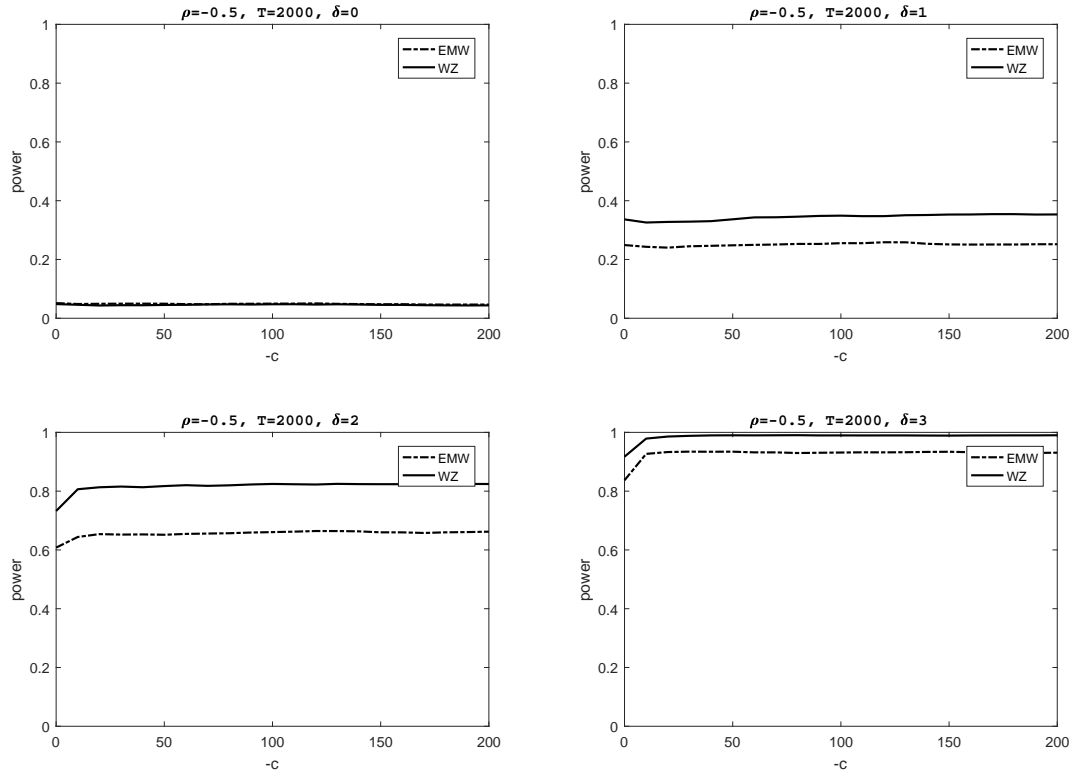


Figure 4.2: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate Gaussian and Power functions of the EMW test, where the true innovation densities f is Multivariate t_3 . The sample size T is 2000.

than the EMW test. Taking the alternative $\delta = 2$ as example, for most of values of c , the power of the EMW test is about 65% while the WZ test attains a power of about 90%.

In Figure 4.2, we keep f unchanged and let g_y and g_x both be Gaussian. Still, both tests are of correct size and again the WZ test is more powerful than the EMW test. However, also observe that the WZ test suffers a loss of power when choosing reference densities that are further away from the true ones (compared to Figure 4.1). When f is Gaussian, the WZ test with Gaussian marginal reference densities shares almost the same size and power as the EMW test (see Figure 4.3). These performances are often related to the traditional Chernoff-Savage property of rank-based inferences: with

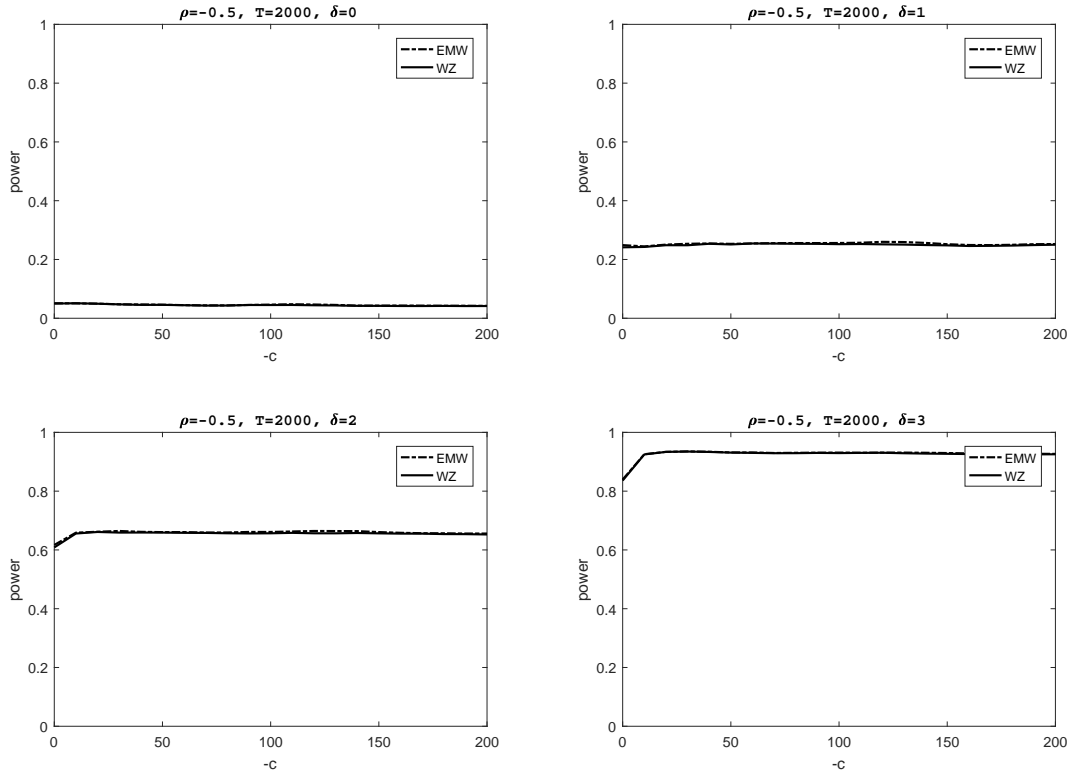


Figure 4.3: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate Gaussian and Power functions of the EMW test, where the true innovation densities f is Multivariate Gaussian. The sample size T is 2000.

fixed reference densities chosen to be Gaussian, the rank-based inference dominates its Gaussian counterpart for any true density f .

However, Figure 4.4 shows that if f is Gaussian and we choose the reference marginal densities that are far from Gaussian (in particular, we choose g_y and g_x both to be univariate t_3 in this example), the WZ test would be less powerful than the EMW test. Nevertheless, the knowledge of f can be revealed with the data at hand.

We also provide some small-sample results for these two tests with $T = 200$ in Figures 4.5–4.8. These three figures can be regarded as the small-sample counterparts of Figures 4.1–4.4, respectively. The conclusions from these figures are very similar: both tests are of decent size (all around 4.5%)

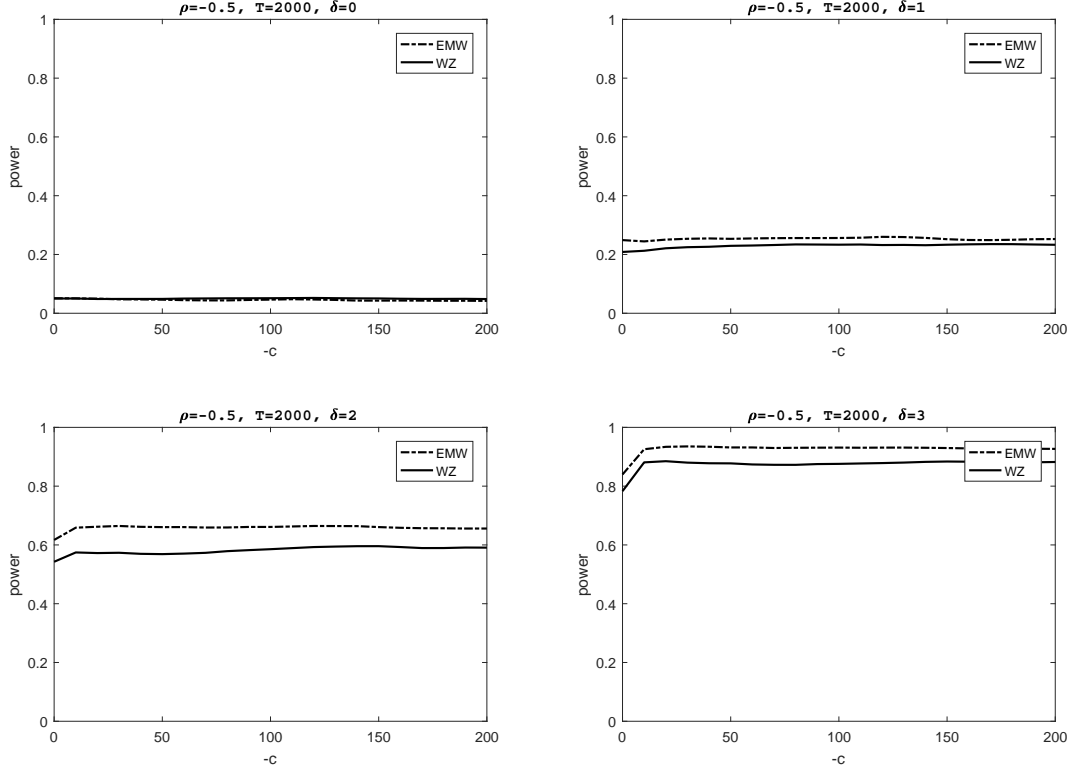


Figure 4.4: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate t_3 and Power functions of the EMW test, where the true innovation densities f is Multivariate Gaussian. The sample size T is 2000.

using the same combinations of $\Lambda_0^{*\epsilon}$ and κ_g for the limit; the WZ test gains considerable power in case of non-Gaussian densities, even if the gain is slightly smaller than for the large-sample cases.

4.6 Conclusions

For the bivariate model with a highly persistent regressor, we propose tests for the regression coefficient that are semiparametric in the sense that the underlying innovation density is regarded as an infinite-dimensional nuisance parameter. Our tests are based on the structural version of the LABF limit experiment of the sequence of experiments, where local alternatives to

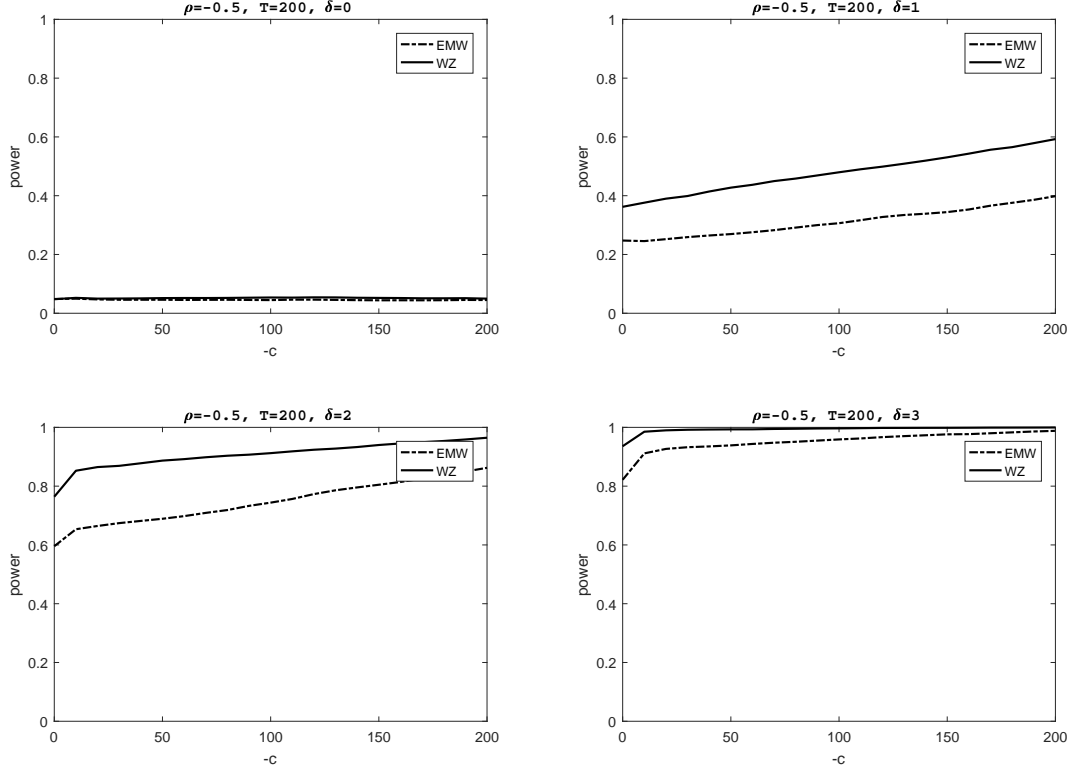


Figure 4.5: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate t_3 and Power functions of the EMW test, where the true innovation densities f is Multivariate t_3 . The sample size T is 200.

the innovation density are modeled nonparametrically using an orthonormal basis.

Specifically, we first derive the maximal invariant in the (structural) limit experiment where the regressor's persistence parameter is assumed to be known. This immediately leads to the semiparametric power envelope for test that are invariant with respect to the innovation density. The associated likelihood ratio thus gives the semiparametric counterparts of the Gaussian sufficient statistics of Jansson and Moreira (2006). To further eliminate the regressor's persistence nuisance parameter, we employ the ALFD approach recently proposed by Elliott, Müller, and Watson (2015).

Subsequently, we propose a class of tests based on the componentwise ranks of the innovations and some chosen marginal reference densities. These

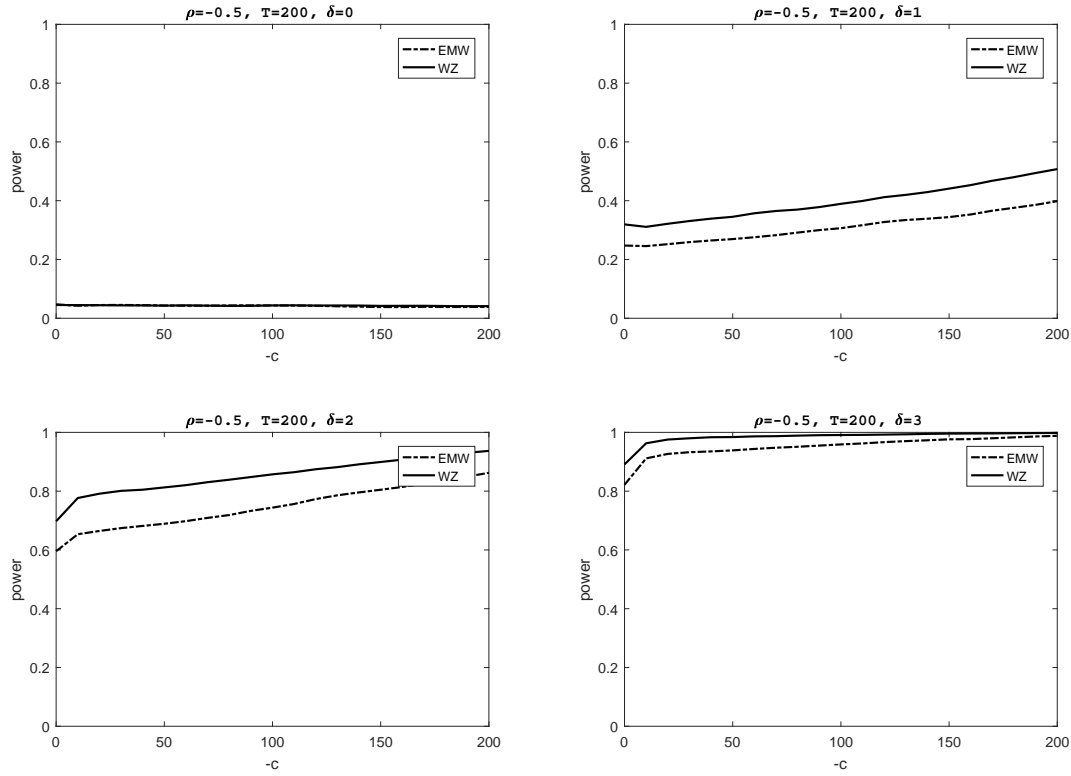


Figure 4.6: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate Gaussian and Power functions of the EMW test, where the true innovation densities f is Multivariate t_3 . The sample size T is 200.

tests are of correct asymptotic size, and have much better power properties than existing tests in the literature that are derived under the assumption of Gaussian innovation densities. Monte Carlo simulations corroborate our asymptotic size and power results and illustrate that the rank-based tests also work well in small samples.

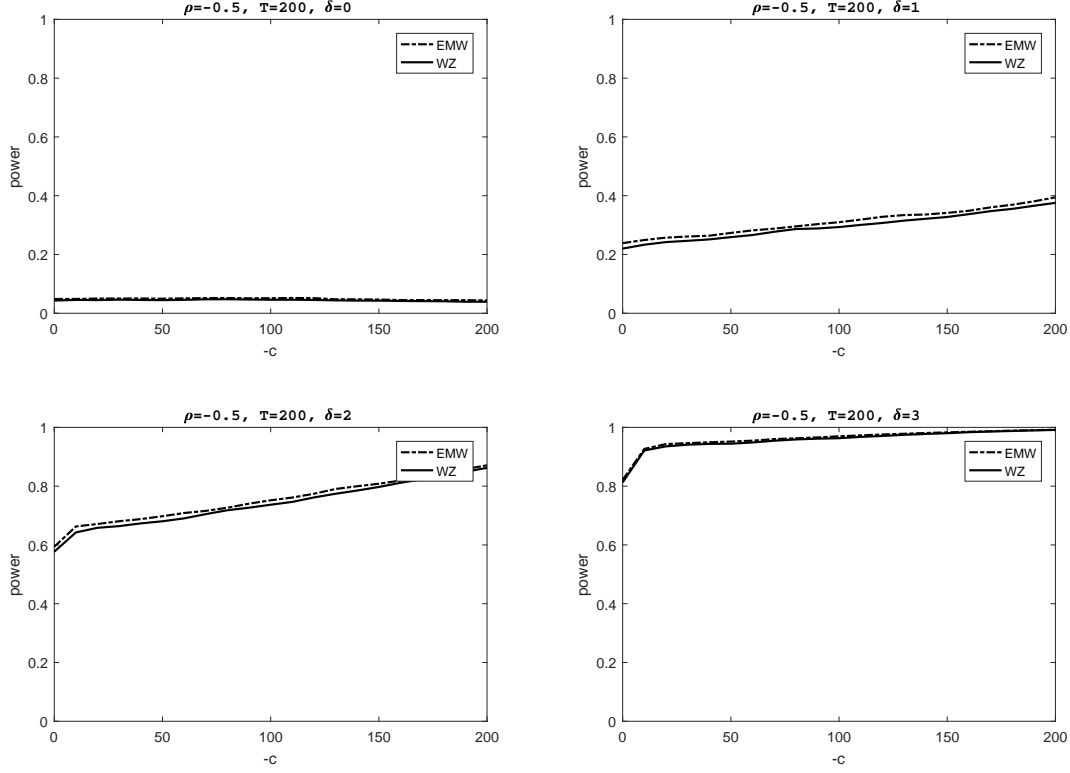


Figure 4.7: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate Gaussian and Power functions of the EMW test, where the true innovation densities f is Multivariate Gaussian. The sample size T is 200.

4.7 Appendix: Proofs

4.7.1 Auxiliaries

The lemma below shows that the partial sum processes introduced in Section 4.2.1 weakly converge to the associated Brownian motions. Due to the i.i.d.-ness of the innovations, the lemma follows, e.g., from the functional central limit theorem VIII. 3.33 in Jacod and Shiryaev (2002).

Lemma 4.7.1. *Let $f \in \mathfrak{F}$ and let, with $m \geq 4$, $k_1, \dots, k_{m-3} \in \mathbb{N}$. Define, with the notation of Section 4.2.1,*

$$\mathcal{W}^{(T)} = (W_{\varepsilon}^{(T)}, W_{\ell_{fy}}^{(T)}, W_{\ell_{fx}}^{(T)}, W_{h_1}^{(T)}, \dots, W_{h_{m-3}}^{(T)})'$$

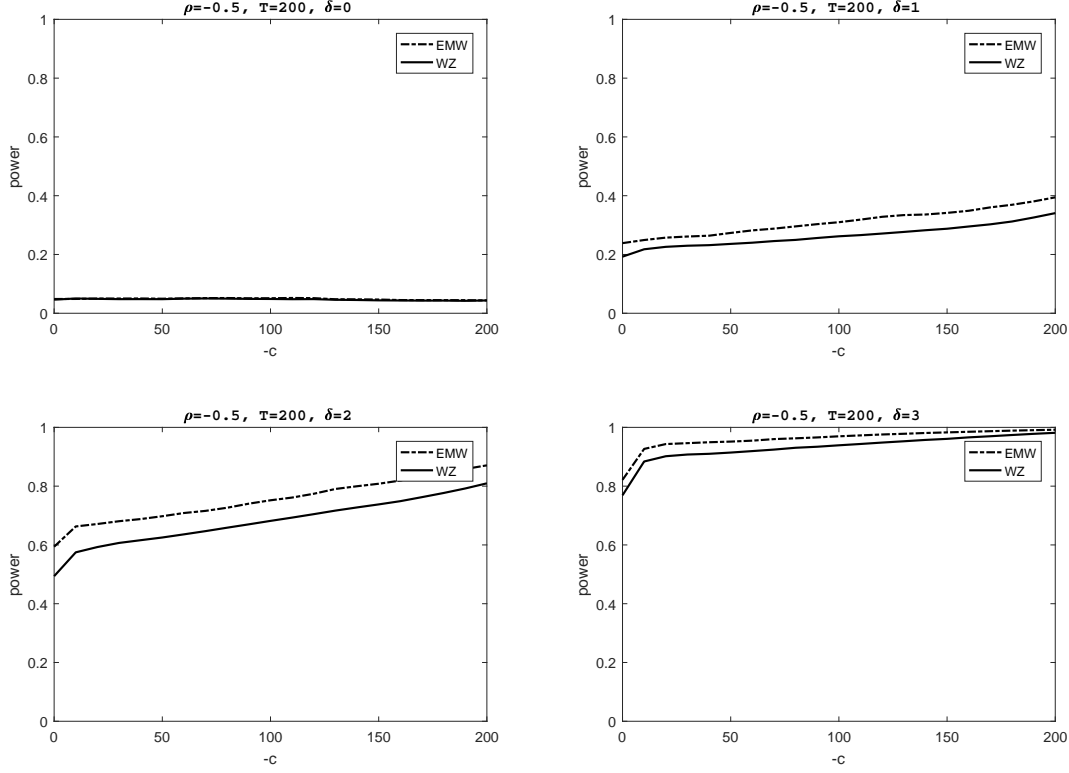


Figure 4.8: Power functions of the WZ Test with reference marginal densities g_y and g_x are univariate t_3 and Power functions of the EMW test, where the true innovation densities f is Multivariate Gaussian. The sample size T is 200.

and

$$\mathcal{W} = (W_\varepsilon, W_{\ell_{f_y}}, W_{\ell_{f_x}}, W_{h_1}, \dots, W_{h_{m-3}})'$$

Then, in $D_{\mathbb{R}^m}[0, 1]$ under $P_{0,0,0,0,f}^{(T)}$, we have

$$\mathcal{W}^{(T)} \Rightarrow \mathcal{W}, \quad (4.7.1)$$

$$\begin{aligned} \langle \mathcal{W}^{(T)}, \mathcal{W}^{(T)} \rangle(1) &= [\mathcal{W}^{(T)}, \mathcal{W}^{(T)}](1) + o_P(1) \\ &= \text{Var}(\mathcal{W}(1)) + o_P(1). \end{aligned} \quad (4.7.2)$$

4.7.2 Main Proofs

Proof of Proposition 4.3.1.

Proof of Part (i):

Suppose $y_t, x_{t-1}, t = 1, 2, \dots, T$, are generated from the model (4.2.1) and (4.2.2). Then the log-likelihood ratio $\log dP_{m,b,c,\eta;f}^{(T)} / dP_{0,0,0;f}^{(T)}$ equals

$$\sum_{t=2}^T \log \left\{ \frac{f(y_t - \mu - \beta x_{t-1}, x_t - \gamma x_{t-1})}{f(y_t^\diamond, \Delta x_t)} \left[1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k (y_t - \mu - \beta x_{t-1}, x_t - \gamma x_{t-1}) \right] \right\}.$$

Using the local parameter perturbations (4.2.5) and (4.2.6), this log-likelihood ratio can be written

$$\log \frac{dP_{m,b,c,\eta;f}^{(T)}}{dP_{0,0,0;f}^{(T)}} = \text{LLR}_I^{(T)}(m, b, c) + \text{LLR}_{II}^{(T)}(m, b, c, \eta), \quad (4.7.3)$$

where

$$\begin{aligned} \text{LLR}_I^{(T)}(m, b, c) &:= \sum_{t=2}^T \log \frac{f\left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, \Delta x_t - \frac{c}{T}x_{t-1}\right)}{f(y_t^\diamond, \Delta x_t)}, \\ \text{LLR}_{II}^{(T)}(m, b, c, \eta) &:= \sum_{t=2}^T \log \left[1 + \frac{1}{\sqrt{T}} \sum_{k=1}^{\infty} \eta_k b_k \left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, \Delta x_t - \frac{c}{T}x_{t-1} \right) \right]. \end{aligned}$$

We first use Proposition 1 in Hallin et al. (2015) to prove

$$\begin{aligned} &\text{LLR}_I^{(T)}(m, b, c) \\ &= \frac{m}{\sqrt{T}} \sum_{t=2}^T \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) + \frac{b}{T} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) + \frac{c}{T} \sum_{t=2}^T x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t) \\ &\quad - \frac{1}{2} \left[m^2 J_{f_{yy}} + (2mb J_{f_{yy}} + 2mc J_{f_{yx}}) \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} \right. \\ &\quad \left. + (b^2 J_{f_{yy}} + c^2 J_{f_{xx}} + 2bc J_{f_{yx}}) \frac{1}{T^2} \sum_{t=2}^T \frac{x_{t-1}^2}{\sigma_x^2} \right] + o_P(1). \end{aligned} \quad (4.7.4)$$

Assumption 4.2.1(a) implies that the density f is differentiable in quadratic mean, i.e.,

$$\frac{\sqrt{f}(e-w)}{\sqrt{f}(e)} = 1 + \frac{1}{2} [w' \ell_f(e) + r(e, w)], \quad e, w \in \mathbb{R}^2, \quad (4.7.5)$$

where

$$E_f r^2(\varepsilon_t, w) = o(w^2). \quad (4.7.6)$$

In the notation of Hallin et al. (2015), we have

$$\begin{aligned} LR_{Tt} &= \frac{f\left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, \Delta x_t - \frac{c}{T}x_{t-1}\right)}{f(y_t^\diamond, \Delta x_t)}, \\ S_{Tt} &= \left(\frac{1}{\sqrt{T}}\sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t), \frac{1}{T}\frac{x_{t-1}}{\sigma_x}\sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t), \frac{1}{T}x_{t-1}\ell_{f_x}(y_t^\diamond, \Delta x_t)\right)', \\ R_{Tt} &= r(\varepsilon_t, w_{Tt}), \end{aligned}$$

where r is implicitly defined in (4.7.5), $w_{Tt} = \left(-\frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, -\frac{c}{T}x_{t-1}\right)'$, and $h_T = (m, b, c)'$. Thus, (4.7.5) implies

$$LR_{Tt} = \left(1 + \frac{1}{2}(h_T' S_{Tt} + R_{Tt})\right)^2.$$

Next, we show that condition (a), (b), (c), and (d) in Proposition 1 of Hallin et al. (2015) are satisfied.

Condition (a). This is immediate since $h_T = (m, b, c)'$ is a constant vector.

Condition (b). Display (2), $E^{(T)}[S_{Tt}|\mathcal{F}_{T,t-1}] = 0$ with $\mathcal{F}_{T,t-1} = \sigma(\varepsilon_t^y, \varepsilon_t^x : t < s)$, follows immediately from the independence of ε_t and $\mathcal{F}_{T,t-1}$, $E_f[\ell_{f_y}(\varepsilon_t)] = 0$, and $E_f[\ell_{f_x}(\varepsilon_t)] = 0$. The second equation in Display (3) is met as

$$\begin{aligned} J_T &:= \sum_{t=1}^T E^{(T)}[S_{Tt} S_{Tt}' | \mathcal{F}_{T,t-1}] \\ &= \sum_{t=1}^T \begin{pmatrix} \frac{1}{T} J_{f_{yy}} & \frac{1}{T^{3/2}} \frac{x_{t-1}}{\sigma_x} J_{f_{yy}} & \frac{1}{T^{3/2}} \frac{x_{t-1}}{\sigma_x} J_{f_{yx}} \\ \frac{1}{T^{3/2}} \frac{x_{t-1}}{\sigma_x} J_{f_{yy}} & \frac{1}{T^2} \frac{x_{t-1}^2}{\sigma_x^2} J_{f_{yy}} & \frac{1}{T^2} \frac{x_{t-1}^2}{\sigma_x^2} J_{f_{yx}} \\ \frac{1}{T^{3/2}} \frac{x_{t-1}}{\sigma_x} J_{f_{yx}} & \frac{1}{T^2} \frac{x_{t-1}^2}{\sigma_x^2} J_{f_{yx}} & \frac{1}{T^2} \frac{x_{t-1}^2}{\sigma_x^2} J_{f_{xx}} \end{pmatrix} \\ \Rightarrow J &:= \begin{pmatrix} J_{f_{yy}} & J_{f_{yy}} \int_0^1 W_\varepsilon(s) ds & J_{f_{yx}} \int_0^1 W_\varepsilon(s) ds \\ J_{f_{yy}} \int_0^1 W_\varepsilon(s) ds & J_{f_{yy}} \int_0^1 W_\varepsilon^2(s) ds & J_{f_{yx}} \int_0^1 W_\varepsilon^2(s) ds \\ J_{f_{yx}} \int_0^1 W_\varepsilon(s) ds & J_{f_{yx}} \int_0^1 W_\varepsilon^2(s) ds & J_{f_{xx}} \int_0^1 W_\varepsilon^2(s) ds \end{pmatrix} = O_P(1), \end{aligned}$$

where the weak convergence follows from a combination of Lemma 4.7.1, Theorem 2.1 in Hansen (1992), and the continuous mapping theorem. Next we verify the conditional Lindeberg condition (the first equation in Display (3)), which is, for all $\delta > 0$,

$$\sum_{t=2}^T E^{(T)} \left[(h_T' S_{Tt})^2 \mathbb{1}_{\{|h_T' S_{Tt}| > \delta\}} | \mathcal{F}_{T,t-1} \right] = o_P(1).$$

Observe

$$\begin{aligned}
& \sum_{t=2}^T \mathbb{E}^{(T)} \left[(h'_T S_{Tt})^2 \mathbb{1}_{\{|h'_T S_{Tt}| > \delta\}} \middle| \mathcal{F}_{T,t-1} \right] \\
&= \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{m}{\sqrt{T}} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) + \frac{b}{T} \frac{x_{t-1}}{\sigma_x} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) + \frac{c}{T} x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t) \right)^2 \right. \\
&\quad \left. \mathbb{1}_{\{(h'_T S_{Tt})^2 > \delta^2\}} \middle| \mathcal{F}_{T,t-1} \right] \\
&\leq 9 \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{m}{\sqrt{T}} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{9(m \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t))^2 > \delta^2 T\}} \middle| \mathcal{F}_{T,t-1} \right] \\
&\quad + 9 \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{b}{T} \frac{x_{t-1}}{\sigma_x} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{9(b x_{t-1} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t))^2 > \delta^2 T^2 \sigma_x^2\}} \middle| \mathcal{F}_{T,t-1} \right] \\
&\quad + 9 \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{c}{T} x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{9(c x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t))^2 > \delta^2 T^2\}} \middle| \mathcal{F}_{T,t-1} \right].
\end{aligned}$$

To complete the proof, we just need to show separately, for any given $\delta > 0$,

$$\begin{aligned}
& \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{m}{\sqrt{T}} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{3|m \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t)| > \delta \sqrt{T}\}} \middle| \mathcal{F}_{T,t-1} \right] = o_P(1), \\
& \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{b}{T} \frac{x_{t-1}}{\sigma_x} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{3|b x_{t-1} \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t)| > \delta T \sigma_x\}} \middle| \mathcal{F}_{T,t-1} \right] = o_P(1), \\
& \sum_{t=2}^T \mathbb{E}^{(T)} \left[\left(\frac{c}{T} x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t) \right)^2 \mathbb{1}_{\{3|c x_{t-1} \ell_{f_x}(y_t^\diamond, \Delta x_t)| > \delta T\}} \middle| \mathcal{F}_{T,t-1} \right] = o_P(1).
\end{aligned}$$

Using the notation $\zeta(M) = \mathbb{E}_f \left[(b \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t))^2 \mathbb{1}_{\{3|b \sigma_y \ell_{f_y}(y_t^\diamond, \Delta x_t)| > \delta T\}} \right]$, we see, for instance, that the left-hand-side of the second term of the previous display is bounded by

$$\zeta \left(\frac{\delta \sqrt{T}}{\|W_\varepsilon^{(T)}\|_\infty} \right) \int_0^1 \left(W_\varepsilon^{(T)}(u-) \right)^2 du = o_P(1),$$

by a combination of Lemma 4.7.1, the continuous mapping theorem, and $\zeta(M) \rightarrow 0$ as $M \rightarrow \infty$ (dominated convergence). The same strategy works for the other two terms.

Condition (c). This condition consists two asymptotic negligibility properties (the Displays (4) and (5) in Hallin et al. (2015)) of the remainder terms

$R_{Tt} = r(\varepsilon_t, w_{Tt})$. Recall $w_{Tt} = \left(-\frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, -\frac{c}{T}x_{t-1}\right)'$, by (4.7.6), we have

$$TE_f [r^2(\varepsilon_t, w_{Tt})|\mathcal{F}_{T,t-1}] = o_P(1), \quad (4.7.7)$$

which ensures the Display (4): $\sum_{t=2}^T E^{(T)} [R_{Tt}^2|\mathcal{F}_{T,t-1}] = o_P(1)$. Display (5), that is

$$\sum_{t=2}^T \left(1 - E^{(T)} [LR_{Tt}|\mathcal{F}_{T,t-1}]\right) = o_P(1),$$

is trivially met by plugging in $LR_{Tt} = \text{LLR}_I^{(T)}(m, b, c)$ to the left-hand-side which gives zero due to the assumed non-negativity of f .

Condition (d). This condition is satisfied since $x_0 = 0$, so that

$$\begin{aligned} \log LR_{Tt} &= \log \frac{f\left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, \Delta x_t - \frac{c}{T}x_{t-1}\right)}{f(y_t^\diamond, \Delta x_t)} \\ &= \log \frac{f\left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y, \Delta x_t\right)}{f(y_t^\diamond, \Delta x_t)} = o_P(1). \end{aligned}$$

Then, for the second term of the log likelihood ratio, $\text{LLR}_{II}^{(T)}(m, b, c, \eta)$, we prove that it equals

$$\frac{\eta'}{\sqrt{T}} \sum_{t=2}^T \sum_k h_k(y_t^\diamond, \Delta x_t) - \frac{1}{2} \left[2aJ'_{fyh}\eta + \left(2bJ'_{f_yh}\eta + 2cJ'_{fxh}\eta\right) \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} + \eta'\eta \right] + o_P(1).$$

This completes the proof for Part (i). Since we assume that the functions h_k , $k \in \mathbb{N}$, are two times continuously differentiable with bounded derivatives, by a Taylor Series expansion, we have

$$h_k\left(y_t^\diamond - \frac{m}{\sqrt{T}}\sigma_y - \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}, \Delta x_t - \frac{c}{T}x_{t-1}\right) \quad (4.7.8)$$

$$\begin{aligned} &= h_k(y_t^\diamond, \Delta x_t) - \left(\frac{m}{\sqrt{T}}\sigma_y + \frac{b}{T}\frac{\sigma_y}{\sigma_x}x_{t-1}\right) \dot{h}_{k,y}(y_t^\diamond, \Delta x_t) \\ &\quad - \frac{c}{T}x_{t-1} \dot{h}_{k,y}(y_t^\diamond, \Delta x_t) + o_P(1), \end{aligned} \quad (4.7.9)$$

where $\dot{h}_{k,y}$ and $\dot{h}_{k,y}$ are the first order derivatives of h_k with respect to the first and second argument, respectively. In this equality, higher order terms

are omitted since the second order derivatives, denoted by $\ddot{h}_{k,yy}$, $\ddot{h}_{k,yx}$, and $\ddot{h}_{k,xx}$, are bounded, i.e., there exists a real number M , such that $|\ddot{h}_{k,yy}| < M$, $|\ddot{h}_{k,yx}| < M$, and $|\ddot{h}_{k,xx}| < M$. Therefore,

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=2}^T \left(\frac{m}{\sqrt{T}} \sigma_y + \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1} \right)^2 \ddot{h}_{k,yy}(y_t^\diamond, \Delta x_t) \\ & < \frac{1}{\sqrt{T}} \sum_{t=2}^T \left(\frac{m}{\sqrt{T}} \sigma_y + \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1} \right)^2 M \\ & \Rightarrow \frac{1}{\sqrt{T}} \left[m^2 \sigma_y^2 + ab \sigma_y^2 \int_0^1 W_\varepsilon(s) ds + b^2 \sigma_y^2 \int_0^1 W_\varepsilon^2(s) ds \right] M = O_P\left(\frac{1}{\sqrt{T}}\right) = o_P(1), \end{aligned}$$

and similar results hold for other higher order terms of $\ddot{h}_{k,yx}$ and $\ddot{h}_{k,xx}$. Also, using $\log(1+x) = x - \frac{1}{2}x^2 + O(x^3)$, we have

$$\begin{aligned} & \text{LLR}_{II}^{(T)}(m, b, c, \eta) \tag{4.7.10} \\ &= \frac{1}{\sqrt{T}} \sum_{t=2}^T \sum_{k=1}^\infty \eta_k b_k \left(y_t^\diamond - \frac{m}{\sqrt{T}} \sigma_y - \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1}, \Delta x_t - \frac{c}{T} x_{t-1} \right) \\ & \quad - \frac{1}{2} \frac{1}{T} \sum_{t=2}^T \left[\sum_{k=1}^\infty \eta_k b_k \left(y_t^\diamond - \frac{m}{\sqrt{T}} \sigma_y - \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1}, \Delta x_t - \frac{c}{T} x_{t-1} \right) \right]^2 + o_P(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=2}^T \sum_{k=1}^\infty \eta_k \left[h_k(y_t^\diamond, \Delta x_t) - \left(\frac{m}{\sqrt{T}} \sigma_y + \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1} \right) \dot{h}_{k,y}(y_t^\diamond, \Delta x_t) - \frac{c}{T} x_{t-1} \dot{h}_{k,x}(y_t^\diamond, \Delta x_t) \right] \\ & \quad - \frac{1}{2} \frac{1}{T} \sum_{t=2}^T \left[\sum_{k=1}^\infty \eta_k b_k \left(y_t^\diamond - \frac{m}{\sqrt{T}} \sigma_y - \frac{b}{T} \frac{\sigma_y}{\sigma_x} x_{t-1}, \Delta x_t - \frac{c}{T} x_{t-1} \right) \right]^2 + o_P(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=2}^T \sum_{k=1}^\infty \eta_k \left[h_k(y_t^\diamond, \Delta x_t) - \left(\frac{m}{\sqrt{T}} + \frac{b}{T} \frac{x_{t-1}}{\sigma_x} \right) J_{f_y h_k} - \frac{c}{T} \frac{x_{t-1}}{\sigma_x} J_{f_x h_k} \right] - \frac{1}{2} \sum_{k=1}^\infty \eta_k^2 + o_P(1) \\ &= \frac{\eta'}{\sqrt{T}} \sum_{t=2}^T \sum_k h_k(y_t^\diamond, \Delta x_t) - a J'_{f_y h} \eta - \left(b J'_{f_y h} \eta + c J'_{f_x h} \eta \right) \frac{1}{T^{3/2}} \sum_{t=2}^T \frac{x_{t-1}}{\sigma_x} - \frac{1}{2} \eta' \eta + o_P(1). \end{aligned}$$

The third equality follows from Lemma 4.7.1, the facts that $E_f[\dot{h}_{k,y}(y_t^\diamond, \Delta x_t)] = \int_{\mathbb{R}^2} \dot{h}_{k,y}(e) f(e) de = h_k(e) f(e)|_{\mathbb{R}^2} - \int_{\mathbb{R}^2} h_k(e) \frac{\dot{f}_y}{f}(e) de = J_{f_y h_k}$ and $E_f[\dot{h}_{k,x}(y_t^\diamond, \Delta x_t)] = J_{f_x h_k}$ by a similar algebra, and the assumption that $E_f[h_k^2(e)] = 1$, and $E_f[h_i(e)h_j(e)] = 0$ when $i \neq j$.

Putting together (4.7.4) and (4.7.10) completes the proof of the LAQ result in Part (i).

Proof of Part (ii): The proof for this part follows immediately from the Functional Central Limit Theorem, see Lemma 4.7.1 and Theorem 2.4 in Chan and Wei (1988).

Proof of Part (iii): Taking the expectation of $\exp(\mathcal{L}(m, b, c, \eta))$ under $\mathbb{P}_{0,0,0,0}$ will directly lead to the result. \square

Proof of Theorem 4.3.2. The proof follows from the definition of the *maximal invariant* in Section 6.2 of Lehmann and Romano (2005), which, in terms of the present problem, is: \mathcal{M} is called maximal invariant with respect to $\mathfrak{G}_{m,\eta}$ if (i) it is invariant, and if (ii) an equality $M(W_\varepsilon, W_h) = M(\widetilde{W}_\varepsilon, \widetilde{W}_h)$, with the mapping M defined in Section 4.3.2, implies that (W_ε, W_h) can be transformed to $(\widetilde{W}_\varepsilon, \widetilde{W}_h)$ with some transformation $\mathbf{g}_{m,\eta} \in \mathfrak{G}_{m,\eta}$. Since (i) is trivially met, the proof is complete if we can show that condition (ii) holds.

Suppose $M(W_\varepsilon(s), W_h(s)) = M(\widetilde{W}_\varepsilon(s), \widetilde{W}_h(s))$, $s \in [0, 1]$. Then

$$W_\varepsilon(s) = \widetilde{W}_\varepsilon(s) \quad \text{and} \quad B_h(s) = \widetilde{B}_h(s).$$

This in turn implies, for $s \in [0, 1]$,

$$W_\varepsilon(s) - \widetilde{W}_\varepsilon(s) = 0 \quad \text{and} \quad W_h(s) - \widetilde{W}_h(s) = c_g s$$

with $c_g = W_h(1) - \widetilde{W}_h(1) \in \mathbb{R}$. This shows that (W_ε, W_h) can indeed be transformed to $(\widetilde{W}_\varepsilon, \widetilde{W}_h)$ by the transformation $\mathbf{g}_{m,\eta} \in \mathfrak{G}_{m,\eta}$ with $mJ_{fyh} + \eta = c_g$. Thus condition (ii) is verified and the proof is complete. \square

Proof of Theorem 4.3.3. Observe that we can decompose the central sequence $\Delta(m, b, c, \eta)$ in (4.3.1) as

$$\Delta(m, b, c, \eta) = \Delta_{\mathcal{M}}(b, c) + \Delta_{\perp}(m, b, c, \eta), \quad (4.7.11)$$

with

$$\Delta_{\perp}(m, b, c, \eta) = mW_{\ell_{fy}}(1) + (bW_{\ell_{fy}}(1) + c[W_{\ell_{fx}}(1) - W_\varepsilon(1)])\overline{W_\varepsilon} + \eta'W_h(1). \quad (4.7.12)$$

Under $\mathbb{P}_{0,0,0,0}$, in view of (4.2.15), $W_h(1)$ is independent of the processes W_ε and B_h . Similarly, $W_{\ell_{fy}}(1)$ and $W_{\ell_{fx}}(1) - W_\varepsilon(1)$ are independent of W_ε and

B_h . Together, this implies that Δ_\perp is independent of \mathcal{M} . As $\Delta_{\mathcal{M}}$ and \mathcal{Q} are obviously \mathcal{M} -measurable, the result follows from

$$\begin{aligned} \mathbb{E} \left[\frac{d\mathbb{P}_{m,b,c,\eta}}{d\mathbb{P}_{0,0,0,0}} | \mathcal{M} \right] &= \mathbb{E} \left[\exp \left(\Delta_{\mathcal{M}}(b, c) + \Delta_\perp(m, b, c, \eta) - \frac{1}{2} \mathcal{Q}(m, b, c, \eta) \right) | \mathcal{M} \right] \\ &= \exp \left(\Delta_{\mathcal{M}}(b, c) - \frac{1}{2} \mathcal{Q}(m, b, c, \eta) \right) \mathbb{E} [\exp \Delta_\perp(m, b, c, \eta) | \mathcal{M}] \\ &= \exp \left(\Delta_{\mathcal{M}}(b, c) - \frac{1}{2} \mathcal{Q}_{\mathcal{M}}(b, c) \right). \end{aligned}$$

This completes the proof. \square

Proof of Proposition 4.3.2. The proposition is nonstandard as it deals with bivariate component-wise ranks. We now have that the so-called Hájek Representation Theorem holds, which is,

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{\lfloor sT \rfloor} \frac{-\dot{g}_y}{g_y} \left[G_y^{-1} \left(\frac{R_{y,t}}{T+1} \right) \right] \\ &= \frac{1}{T} \sum_{t=1}^{\lfloor sT \rfloor} \frac{-\dot{g}_y}{g_y} [G_y^{-1}(F_y(\varepsilon_t^y))] - \frac{1}{T} \sum_{t=1}^T \frac{-\dot{g}_y}{g_y} [G_y^{-1}(F_y(\varepsilon_t^y))] + o_P(1). \end{aligned}$$

The equivalent result holds for the ranks $R_{x,t}$, with y replaced by x in the above expression. The claim now follows from the functional central limit theorem applied to $\frac{-\dot{g}_y}{g_y} [G_y^{-1}(F_y(\varepsilon_t^y))]$ and $\frac{-\dot{g}_x}{g_x} [G_x^{-1}(F_x(\varepsilon_t^x))]$, jointly with $W_\varepsilon^{(T)}$ and $W_{\ell_f}^{(T)}$. \square

Proof of Corollary 4.3.3. The behavior of W_ε under $\mathbb{P}_{m,b,c,\eta}$ is already given in the structural limit experiment associated to the maximal invariant \mathcal{M} in Corollary 4.3.2. To get the behavior of B_{ℓ_g} under $\mathbb{P}_{m,b,c,\eta}$, first decompose it as

$$B_{\ell_g}(s) = vB_\varepsilon(s) + AB_{\ell_f}(s) + B_\perp(s)$$

for some $v \in \mathbb{R}^{2 \times 1}$ and $A \in \mathbb{R}^{2 \times 2}$, where B_\perp is the Brownian bridge of a Brownian motion W_\perp which is independent of W_ε and W_{ℓ_f} . The values of v and A satisfy the relation

$$J_{gf} = \text{Cov}(W_{\ell_g}(1), W_{\ell_f}(1))$$

$$\begin{aligned}
&= \text{Cov}(vW_\varepsilon(1) + AW_{\ell_f}(1) + W_\perp(1), W_{\ell_f}(1)) \\
&= v\mathbf{e}'_1 + AJ_f.
\end{aligned}$$

Then, under $\mathbb{P}_{m,b,c,\eta}$, we have

$$\begin{aligned}
dB_{\ell_g}(s) &= vdB_\varepsilon(s) + AdB_{\ell_f}(s) + dB_\perp(s) \\
&= v \left[cW_\varepsilon^\mu ds + dB^{Z_\varepsilon}(s) \right] + A \left[J_f(b, c)'W_\varepsilon^\mu(s)ds + dB^{Z_{\ell_f}}(s) \right] + dB_\perp(s) \\
&= (v\mathbf{e}'_1 + AJ_f)(b, c)'W_\varepsilon^\mu(s)ds + \left[dB^{Z_\varepsilon}(s) + dB^{Z_{\ell_f}}(s) + dB^{Z_\perp}(s) \right] \\
&= J_{gf}(b, c)'W_\varepsilon^\mu(s)ds + dB^{Z_{\ell_g}}(s).
\end{aligned}$$

□

4.8 Appendix: Switching Tests to Standard Case

The numerical approach of Elliott, Müller, and Watson (2015) needs to discretize the nuisance parameter space under the null hypothesis (and the associated mesh is regarded as the support of $\Lambda_0^{*\epsilon}$). However, in the present case, the null parameter space of c is $(-\infty, 0]$, which is unbounded. This complicates the algorithm in terms of computation. To address this issue, Elliott, Müller, and Watson (2015) proposes to switch to a standard test when $|c|$ is large enough so that the predictor essentially behaves like a stationary time series. In that case, the problem reduces to a standard regression test with a stationary regressor. In particular, the authors propose to use a “switching” function $\chi = \mathbb{1}\{\hat{c} < K\}$ based on some estimator \hat{c} of parameter c and a chosen “threshold” K to distinguish the nonstandard situation from the standard one. Then, one can employ the following “combined” test function

$$\varphi_{n,s,\chi}(S) = \chi\varphi_s(S) + (1 - \chi)\varphi_n(S), \quad (4.8.1)$$

where φ_s is some test for the standard case, and φ_n is the test (4.4.12) for the nonstandard case. For the standard test φ_s , following the argument in the same paper, we use the semiparametric version of the maximal likelihood

t -test

$$\varphi_s(S) = \mathbb{1} \{b^*/\sigma_{b^*} > \kappa_s\} \quad (4.8.2)$$

with

$$b^* = \frac{S_1}{S_3 J_{f_{yy}}} - \frac{J_{f_{yx}}}{J_{f_{yy}}} c^*, \quad c^* = \frac{S_2 - (J_{f_{yx}}/J_{f_{yy}})S_1}{[(J_{f_{xx}} - 1) - J_{f_{yx}}^2/J_{f_{yy}}]S_3 + S_4}, \quad \text{and}$$

$$\sigma_{b^*} = \sqrt{\frac{1}{J_{f_{yy}}S_3} + \left(\frac{J_{f_{yx}}}{J_{f_{yy}}}\right)^2 \frac{1}{[(J_{f_{xx}} - 1) - J_{f_{yx}}^2/J_{f_{yy}}]S_3 + S_4}}.$$

Here b^* and c^* are actually the maximum likelihood estimators of b and c based on the likelihood ratio $\mathcal{L}_{\mathcal{M}}(b, c)$ in (4.4.1).

The proof of the following lemma can be found in the Supplementary Material of Elliott, Müller, and Watson (2015) (Appendix C.4).

Lemma 4.8.1. *For $s \in [0, 1]$, let $Z_1(s)$ and $Z_2(s)$ be two independent standard Brownian motions, and $W_1(s)$ be the associated Ornstein-Uhlenbeck process of $Z_1(s)$, defined by $dW_1(s) = cW_1(s)ds + dZ_1(s)$. Define the de-meaned process $W_1^\mu(s) = W_1(s) - \int_0^1 W_1(s)ds$. Then, as $c \rightarrow -\infty$, we have*

$$\begin{pmatrix} \sqrt{-2c} \int_0^1 W_1(s) dZ_1(s) \\ \sqrt{-2c} \int_0^1 W_1^\mu(s) dZ_2(s) \\ -2c \int_0^1 (W_1(s))^2 ds \\ -2c \int_0^1 (W_1^\mu(s))^2 ds \end{pmatrix} \Rightarrow \begin{pmatrix} z_1 \\ z_2 \\ 1 \\ 1 \end{pmatrix}, \quad (4.8.3)$$

where z_1 and z_2 are two independent standard normal random variables.

Lemma 4.8.2. *Suppose the sufficient statistics S_1, S_2, S_3, S_4 are defined in (4.4.2), where the behavior of $(W_\varepsilon, B_{\ell_{fy}}, B_{\ell_{fx}})'$ are described by the limit experiment $\mathcal{E}_{\mathcal{M}}(f)$ in Corollary 4.3.2. Then, under $\mathbb{P}_{m,0,c,\eta}$ and as $c \rightarrow -\infty$, we have*

$$\sqrt{-2c} \left(\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} J_{f_{yx}} \\ J_{f_{xx}} \end{pmatrix} \right) \Rightarrow \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J_{f_{yy}} & J_{f_{yx}} \\ J_{f_{yx}} & J_{f_{xx}} \end{pmatrix} \right], \quad (4.8.4)$$

$$-2cS_3 \Rightarrow 1, \quad \text{and} \quad -2cS_4 \Rightarrow 1. \quad (4.8.5)$$

Subsequently, still under $\mathbb{P}_{m,0,c,\eta}$ and as $c \rightarrow -\infty$, we have

$$b^*/\sigma_{b^*} \Rightarrow \mathcal{N}(0, 1). \quad (4.8.6)$$

Proof of Lemma 4.8.2. Note that in this proof, all convergence results (as $c \rightarrow -\infty$) come immediately from Lemma 4.8.1.

First, we give the convergence results of statistics S_3 and S_4 : Recall $dW_\varepsilon(s) = cW_\varepsilon(s)ds + dZ_\varepsilon(s)$ for $s \in [0, 1]$ which makes $W_\varepsilon(s)$ an Ornstein-Uhlenbeck process. Then we have, as $c \rightarrow -\infty$,

$$\begin{aligned} -2cS_3 &= -2c \left(\overline{W_\varepsilon^2} - (\overline{W_\varepsilon})^2 \right) = -2c \int_0^1 (W_\varepsilon^\mu(s))^2 ds \Rightarrow 1, \\ -2cS_4 &= -2c \overline{W_\varepsilon^2} = -2c \int_0^1 (W_\varepsilon(s))^2 ds \Rightarrow 1. \end{aligned} \quad (4.8.7)$$

Next, we give the convergence results of statistics S_1 and S_2 : To this end, we state beforehand some derivative results of Lemma 4.8.1: Define $W_\varepsilon^\mu(s) = W_\varepsilon(s) - \int_0^1 W_\varepsilon(s)ds$ for $s \in [0, 1]$ and any infinite-dimensional vector $A_1, A_2 \in \mathbb{R}^{\infty \times 1}$, we have

$$\begin{pmatrix} -2cA_1' \int_0^1 W_\varepsilon^\mu(s) dZ_h(s) \\ -2cA_2' \int_0^1 W_\varepsilon^\mu(s) dZ_h(s) \end{pmatrix} \Rightarrow \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A_1' A_1 & A_1' A_2 \\ A_1' A_2 & A_2' A_2 \end{pmatrix} \right].$$

Hence, decompose $\sqrt{-2c}S_1$ as

$$\begin{aligned} \sqrt{-2c}S_1 &= \sqrt{-2c} \int_0^1 W_\varepsilon(s) dB_{\ell_{fy}}(s) \\ &= \sqrt{-2c} \int_0^1 W_\varepsilon^\mu(s) dW_{\ell_{fy}}(s) \\ &= \sqrt{-2c} \int_0^1 W_\varepsilon^\mu(s) dZ_{\ell_{fy}}(s) + \sqrt{-2c} J_{f_{yx}} \int_0^1 W_\varepsilon^\mu(s) W_\varepsilon(s) ds \\ &= \sqrt{-2c} \int_0^1 W_\varepsilon^\mu(s) dZ_{\ell_{fy}}(s) - \frac{\sqrt{-2c}}{2} J_{f_{yx}} \left(-2c \int_0^1 (W_\varepsilon^\mu(s))^2 ds \right), \end{aligned}$$

then we have

$$\sqrt{-2c}S_1 + \frac{\sqrt{-2c}}{2} J_{f_{yx}} \Rightarrow \mathcal{N}(0, J_{f_{yy}}).$$

Similarly, decompose $\sqrt{-2c}S_2$ as

$$\begin{aligned} \sqrt{-2c}S_2 &= \sqrt{-2c} \left(\int_0^1 W_\varepsilon(s) dB_{\ell_{fx}}(s) + W_\varepsilon(1) \int_0^1 W_\varepsilon(s) ds \right) \\ &= \sqrt{-2c} \left(\int_0^1 W_\varepsilon(s) dW_\varepsilon(s) + J_{f_{xh}} \int_0^1 W_\varepsilon^\mu(s) dW_h(s) \right) \end{aligned}$$

$$\begin{aligned}
&= \sqrt{-2c} \left(\int_0^1 W_\varepsilon(s) dZ_\varepsilon(s) + J_{f_x h} \int_0^1 W_\varepsilon^\mu(s) dZ_h(s) \right) \\
&\quad - \frac{\sqrt{-2c}}{2} \left(-2c \int_0^1 W_\varepsilon(s)^2 ds - 2c J_{f_x h} J'_{f_x h} \int_0^1 (W_\varepsilon^\mu(s))^2 ds \right),
\end{aligned}$$

then, together with the fact that $J_{f_{xx}} = 1 + J_{f_x h} J'_{f_x h}$, we have

$$\sqrt{-2c} S_2 + \frac{\sqrt{-2c}}{2} J_{f_{xx}} \Rightarrow \mathcal{N}(0, J_{f_{xx}}).$$

Moreover, since the covariance of $\sqrt{-2c} S_1$ and $\sqrt{-2c} S_1$ is $J_{f_y h} J'_{f_x h} = J_{f_{yx}}$, it then follows that

$$\sqrt{-2c} \left(\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} J_{f_{yx}} \\ J_{f_{xx}} \end{pmatrix} \right) \Rightarrow \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J_{f_{yy}} & J_{f_{yx}} \\ J_{f_{yx}} & J_{f_{xx}} \end{pmatrix} \right] \quad (4.8.8)$$

Finally, we give the convergence result that $b^*/\sigma_{b^*} \Rightarrow \mathcal{N}(0, 1)$: By (4.8.8), we can easily show that

$$\sqrt{-2c} \left(\begin{pmatrix} S_1 \\ S_2 - \frac{J_{f_{yx}}}{J_{f_{yy}}} S_1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} J_{f_{yx}} \\ J_{f_{xx}} - \frac{J_{f_{yx}}^2}{J_{f_{yy}}} \end{pmatrix} \right) \Rightarrow \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J_{f_{yy}} & 0 \\ 0 & J_{f_{xx}} - \frac{J_{f_{yx}}^2}{J_{f_{yy}}} \end{pmatrix} \right].$$

Thus, after some algebra, we have

$$\begin{aligned}
\frac{b^*}{\sqrt{-2c}} &= \frac{\sqrt{-2c} S_1}{-2c S_3 J_{f_{yy}}} - \frac{J_{f_{yx}}}{J_{f_{yy}}} \frac{c^*}{\sqrt{-2c}} \\
&= \frac{\sqrt{-2c} S_1}{-2c S_3 J_{f_{yy}}} - \frac{J_{f_{yx}}}{J_{f_{yy}}} \frac{\sqrt{-2c} S_2 - (J_{f_{yx}}/J_{f_{yy}}) \sqrt{-2c} S_1}{[(J_{f_{xx}} - 1) - J_{f_{yx}}^2/J_{f_{yy}}](-2c S_3) + (-2c S_4)} \\
&\Rightarrow \mathcal{N} \left(0, \left(\frac{1}{J_{f_{yy}}} + \left(\frac{J_{f_{yx}}}{J_{f_{yy}}} \right)^2 \frac{1}{J_{f_{xx}} - J_{f_{yx}}^2/J_{f_{yy}}} \right) \right).
\end{aligned}$$

Moreover, following (4.8.7), we have

$$\begin{aligned}
\frac{\sigma_{b^*}}{\sqrt{-2c}} &= \sqrt{\frac{1}{J_{f_{yy}} S_3} + \left(\frac{J_{f_{yx}}}{J_{f_{yy}}} \right)^2 \frac{1}{[(J_{f_{xx}} - 1) - J_{f_{yx}}^2/J_{f_{yy}}](-2c S_3) + (-2c S_4)}} \\
&\Rightarrow \sqrt{\frac{1}{J_{f_{yy}}} + \left(\frac{J_{f_{yx}}}{J_{f_{yy}}} \right)^2 \frac{1}{J_{f_{xx}} - J_{f_{yx}}^2/J_{f_{yy}}}},
\end{aligned}$$

which completes the proof. \square

For the standard part, mimicking the construction of standard test φ_s in (4.8.2), we define

$$b_g^* = \frac{S_1^g}{S_3^g J_{p_{yy}}} - \frac{J_{p_{yx}}}{J_{p_{yy}}} c_g^*, \quad c_g^* = \frac{S_2^g - (J_{p_{yx}}/J_{p_{yy}})S_1^g}{[(J_{p_{xx}} - 1) - J_{p_{yx}}^2/J_{p_{yy}}]S_3^g + S_4^g},$$

and

$$\sigma_{b_g^*} = \sqrt{\frac{1}{J_{p_{yy}}S_3^g} + \left(\frac{J_{p_{yx}}}{J_{p_{yy}}}\right)^2 \frac{1}{[(J_{p_{xx}} - 1) - J_{p_{yx}}^2/J_{p_{yy}}]S_3^g + S_4^g}}.$$

The following lemma can be regarded as the rank-based version of Lemma 4.8.2.

Lemma 4.8.3. *Suppose S^g are defined in (4.4.8)-(4.4.9), where the behavior of $(W_\varepsilon, B_{\ell_{gy}}, B_{\ell_{gx}})'$ are described in Corollary 4.3.3. Then, under H_0 and as $c \rightarrow -\infty$, we have*

$$b_g^*/\sigma_{b_g^*} \Rightarrow \mathcal{N}(0, 1).$$

Proof of Lemma 4.8.3. When $c \rightarrow -\infty$, recall that $-2cS_3 \rightarrow 1$ and $-2cS_4 \rightarrow 1$, then we have

$$\begin{aligned} \frac{\sigma_{b_g^*}}{\sqrt{-2c}} &= \sqrt{\frac{1}{J_{p_{yy}}(-cS_3^g)} + \left(\frac{J_{p_{yx}}}{J_{p_{yy}}}\right)^2 \frac{1}{[(J_{p_{xx}} - 1) - J_{p_{yx}}^2/J_{p_{yy}}](-2cS_3^g) + (-2cS_4^g)}} \\ &\Rightarrow \sqrt{\frac{1}{J_{p_{yy}}} + \left(\frac{J_{p_{yx}}}{J_{p_{yy}}}\right)^2 \frac{1}{J_{p_{xx}} - J_{p_{yx}}^2/J_{p_{yy}}}} \\ &= \sqrt{\frac{1}{J_{g_y}}}, \end{aligned}$$

where the last equality holds by plugging in J_p defined in (4.4.11). Similarly, by the same steps, we have as $c \rightarrow -\infty$

$$\begin{aligned} \frac{b_g^*}{\sqrt{-2c}} &= \frac{S_1^g}{S_3^g J_{p_{yy}}} - \frac{J_{p_{yx}}}{J_{p_{yy}}} \frac{S_2^g - (J_{p_{yx}}/J_{p_{yy}})S_1^g}{[(J_{p_{xx}} - 1) - J_{p_{yx}}^2/J_{p_{yy}}]S_3^g + S_4^g} \\ &= \sqrt{-2c} \left(\frac{S_1^g}{J_{p_{yy}}} - \frac{J_{p_{yx}}}{J_{p_{yy}}} \frac{S_2^g - (J_{p_{yx}}/J_{p_{yy}})S_1^g}{J_{p_{xx}} - J_{p_{yx}}^2/J_{p_{yy}}} \right) \\ &= \frac{\sqrt{-2c}}{J_{g_y}} (S_1^g + \rho S_2^g) \\ &= \frac{\sqrt{-2c}}{J_{g_y}} \int_0^1 W_\varepsilon(s) dB_{g_y}(s) \end{aligned}$$

where $B_{g_y} := B_{\ell_{g_y}} + \rho B_{\ell_{g_x}}$. The last equality holds because $\overline{W_\varepsilon} \rightarrow 0$ as $c \rightarrow -\infty$. It is not hard to find that, based on the construction in (4.3.12)-(4.3.13), B_{g_y} is the limit of the partial-sum process $\frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} -\frac{g_y}{g_y} \left[G_y^{-1} \left(\frac{R_{y,t}}{T+1} \right) \right]$. Thus, under H_0 , B_{g_y} is a Brownian bridge rather than a bridge process of an OU process. As $c \rightarrow -\infty$, by Lemma 4.8.1 and the fact that $\overline{W_\varepsilon} = 0$, we have

$$\frac{\sqrt{-2c}}{J_{g_y}} \int_0^1 W_\varepsilon(s) dB_{g_y}(s) \Rightarrow \mathcal{N} \left(0, \frac{1}{J_{g_y}} \right),$$

which in turn completes the proof. \square

Then we have the feasible standard test

$$\varphi_{g,s}(S^g, \rho) = \mathbb{1} \left\{ b_g^* / \sigma_{b_g^*} > \kappa_{g,s} \right\} \quad (4.8.9)$$

where $\kappa_{g,s}$ is the $(1 - \alpha)$ -quantile of a standard normal distribution.

Similarly, employing the “combined” functional form as in (4.8.1), we have the rank-based test

$$\varphi_{g,\chi_g}(S^g, \rho) = \chi_g \varphi_{g,s}(S^g, \rho) + (1 - \chi_g) \varphi_{g,n}(S^g, \rho), \quad (4.8.10)$$

where $\chi_g = \mathbb{1}\{c_g^* < K_g\}$.

Replace S^g by their finite-sample counterpart \widehat{S}^g , and replace ρ by an appropriate estimate $\widehat{\rho}$ to obtain a feasible test $\varphi_{g,\chi_g}(\widehat{S}^g, \widehat{\rho})$.

Bibliography

- AHN, S.A., S.B. FOTOPOULOS, AND L. HE (2003). Unit Root Tests with Infinite Variance Errors. *Econometric Reviews*, 20, 461–483.
- ANDERSON, T.W., AND H., RUBIN (1949). Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, 21, 570–582.
- ANDREWS, D.W.K., AND V., MARMER (2008). Exactly Distribution-Free Inference in Instrumental Variables Regression with Possibly Weak Instruments. *Journal of Econometrics*, 142, 183–200.
- ANDREWS, D.W.K., M.J. MOREIRA, AND J.H. STOCK (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, 74, 715–752.
- ANDREWS, D. W. K., AND W. PLOBERGER (1994). Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative. *Econometrica*, 62, 1383–1414.
- ANDREWS, D.W.K., AND G., SOARES (2007). Rank Tests for Instrumental Variables Regression with Weak Instruments. *Econometric Theory*, 23, 1033–1082.
- BHARGAVA, A. (1986). On the Theory of Testing for Unit Roots in Observed Time Series. *The Review of Economic Studies*, 53, 369–384.
- BICKEL, P.J. (1982). On Adaptive Estimation. *The Annals of Statistics*, 10, 647–671.

- BICKEL, P.J., C.A.J. KLAASSEN, Y. RITOV, AND J.A. WELLNER (1998), *Efficient and Adaptive Estimation for Semiparametric Models*. Amsterdam: Springer-Verlag.
- BOSWIJK, H.P., M. JANSSON, AND M.Ø. NIELSEN (2015). Improved Likelihood Ratio Tests for Cointegration Rank in the VAR Model. *Journal of Econometrics*, 184, 97–110.
- BREITUNG, J. AND C. GOURIÉROUX (1997). Rank Tests for Unit Roots. *Journal of Econometrics*, 81, 7–27.
- CALLEGARI, F., N. CAPPuccio, AND D. LUBIAN (2003). Asymptotic Inference in Time Series Regressions with a Unit Root and Infinite Variance Errors. *Journal of Statistical Planning and Inference*, 116, 277–303.
- CAMPBELL, B. AND J.M. DUFOUR (1995). Exact Nonparametric orThogonality and Random Walk Tests. *The Review of Economics and Statistics*, 77, 1–16.
- CAMPBELL, B. AND J.M. DUFOUR (1997). Exact Nonparametric Tests of orThogonality and Random Walk in the Presence of a Drift Parameter. *International Economic Review*, 38, 151–173.
- CAMPBELL, J.Y., AND M. YOGO (2005). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics*, 81, 27–60.
- CASSART, D., HALLIN, M., AND PAINDAVEINE, D. (2010). On the Estimation of Cross-Information Quantities in Rank-Based Inference. *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jureckova*, 7, 35–45.
- CATTANEO, M.D., R.K. CRUMP, AND M. JANSSON (2012). Optimal Inference for Instrumental Variables Regression with Non-Gaussian Errors. *Journal of Econometrics*, 167, 1–15.
- CAVANAGH C. L. , G. ELLIOTT, AND J. H. STOCK (1995). Inference in Models with Nearly Integrated Regressors. *Econometric Theory*, 11, 1131–1147.

- CHAN, N.H. AND C.Z. WEI (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *The Annals of Statistics*, 16, 367–401.
- CHERNOFF, H. AND I.R. SAVAGE (1958). Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics. *The Annals of Mathematical Statistics*, 29, 972–994.
- CHOI, I. (2015). *Almost All About Unit Roots*. Cambridge University Press.
- COUDIN, E. AND J.M. DUFOUR (2009). Finite-Sample Distribution-Free Inference in Linear Median Regression under Heteroskedasticity and Non-linear Dependence of Unknown Form. *Econometrics Journal*, 12, 19–49.
- DICKEY, D.A. AND W.A. FULLER (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431.
- DICKEY, D.A. AND W.A. FULLER (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49, 1057–1072.
- DUFOUR, J.-M. AND M.L. KING (1991). Optimal invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Non-stationary AR(1) Errors. *Journal of Econometrics*, 47, 115–143.
- ELLIOTT, G. AND U.K. MÜLLER (2006). Minimizing the Impact of the Initial Condition on Testing for Unit Roots. *Journal of Econometrics*, 135, 285–310.
- ELLIOTT, G., U.K. MÜLLER, and MARK W. WATSON (2015). Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis. *Econometrica*, 83, 771–811.
- ELLIOTT, G., T.J. ROTHENBERG, and J.H. STOCK (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64, 813–836.
- ELLIOTT, G., AND J.H. STOCK (1994). Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown. *Econometric*

- Theory*, 10, 672–700.
- GUSHCHIN, A.A. (1996). Asymptotic Optimality of Parameter Estimators under the LAQ Condition. *Theory of Probability and its Applications*, 40, 261–272.
- HALL, P. AND C.C. HEYDE (1980). *Martingale Limit Theory and its Application*. Academic Press, New York.
- HALLIN, M., R. VAN DEN AKKER, and B.J.M. WERKER (2011). A Class of Simple Distribution-Free Rank-Based Unit Root Tests. *Journal of Econometrics*, 163, 200–214.
- HALLIN, M., and B.J.M. WERKER (2003). Semiparametric Efficiency, Distribution-Freeness and Invariance, *Bernoulli*, 9(1), 137–165.
- HALLIN, M., R. VAN DEN AKKER, and B.J.M. WERKER (2015). On Quadratic Expansions of Log-Likelihoods and a General Asymptotic Linearity Result. *Mathematical Statistics and Limit Theorems : Festschrift in honour of Paul Deheuvels*, 12 p.
- HALLIN, M., R. VAN DEN AKKER, AND B.J.M. WERKER (2016). Semiparametric Error-Correction Models for Cointegration with Trends: Pseudo-Gaussian and Optimal Rank-Based Tests of the Cointegration Rank. *Journal of Econometrics*, 190, 46–61.
- HANSEN, B.E. (1992). Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory*, 8, 489–500.
- HASAN, M.N. (2001). Rank Tests of Unit Root Hypothesis with Infinite Variance Errors. *Journal of Econometrics*, 104, 49–65.
- HASAN, M.N., AND R.W. KOENKER (1997). Robust Rank Tests of the Unit Root Hypothesis. *Econometrica*, 65, 133–161.
- JACOD, J. AND A.N. SHIRYAEV (2002). *Limit Theorems for Stochastic Processes*. Berlin: Springer.
- JANSSON, M. (2008). Semiparametric Power Envelopes for Tests of the Unit

- Root Hypothesis. *Econometrica*, 76, 1103–1142.
- JANSSON, M., AND MARCELO J. MOREIRA (2006). Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica*, 74, 681–714.
- JANSSON, M. AND NIELSEN, M. (2012). Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis. *Econometrica*, 80, 2321–2332.
- JEGANATHAN, P. (1995). Some Aspects of Asymptotic Theory with Applications to Time Series Models. *Econometric Theory*, 11, 818–887.
- (1997). On Asymptotic Inference in Linear Cointegrated Time Series Systems. *Econometric Theory*, 13, 692–745.
- JOHANSEN, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59, 1551–1580.
- KAGANA, A. AND Z. LANDSMANB (1999). Relation between the Covariance and Fisher Information Matrices. *Statistics and Probability Letters*, 42, 7–13.
- KARATZAS, I. AND S.E. SHREVE (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, second edition.
- KLAASSEN, C. A. J. (1987). Consistent Estimation of the Influence Function of Locally Asymptotically Linear Estimators. *Annals of Statistics*, 15, 1548–1562.
- KLEIBERGEN, F., (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70, 1781–1803.
- LE CAM, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.
- LE CAM, L.M., G.L. YANG (2000). *Asymptotics in Statistics Some Basic Concepts* (Second Edition). Springer-Verlag, New York.
- LEHMANN, ERICH L., ROMANO, JOSEPH P. (2005). *Testing Statistical*

- Hypothesis* (Third Edition). Springer, New York.
- MAYER-WOLF, E (1990). The Cramer-Rao Functional and Limiting Laws. *The Annals of Probability*, 18, 840–850.
- MOREIRA, M.J., (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71, 1027–1048.
- MÜLLER, U.K. (2011). Efficient Tests under a Weak Convergence Assumption. *Econometrica*, 79, 395–435.
- MÜLLER, U.K., AND G. ELLIOTT (2003). Tests for Unit Roots and the Initial Condition. *Econometrica*, 71, 1269–1286.
- MÜLLER, U.K., AND ANDRIY NORETS (2016). Credibility of Confidence Sets in Nonstandard Econometric Problems. *Econometrica*, 84, 2183–2213.
- MÜLLER, U.K., AND M. W. WATSON (2008). Testing Models of Low-Frequency Variability. *Econometrica*, 76, 979–1016.
- NEWBY, W.K., (1990). Semiparametric Efficient Bounds. *Journal of Applied Econometrics*, 5, 99–135.
- PATTERSON, K. (2011). *Unit Root Tests in Time Series, Volume 1: key concepts and problems*. Palgrave Macmillan.
- PATTERSON, K. (2012). *Unit Root Tests in Time Series, Volume 2: extensions and developments*. Palgrave Macmillan.
- PHILLIPS, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica*, 55, 277–301.
- PHILLIPS, P.C.B. (1988). Regression Theory for Near-Integrated Time Series. *Econometrica*, 56, 1021–1043.
- PHILLIPS, P.C.B. (2014). On Confidence Intervals for Autoregressive Roots and Predictive Regression. *Econometrica*, 82, 1177–1195.
- PHILLIPS, P.C.B., AND P. PERRON (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75, 335–346.

- PLOBERGER, W. (2004). Complete Class of Tests when the Likelihood is Locally Asymptotically Quadratic. *Journal of Econometrics*, 118, 67–94.
- PLOBERGER, W. (2008). Admissible and Nonadmissible Tests in Unit-Root-Like Situations. *Econometric Theory*, 24, 15–42.
- ROTHENBERG, T.J., AND J.H. STOCK (1997). Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations. *Journal of Econometrics*, 80, 268–286.
- SAIKKONEN, P. AND R. LUUKKONEN (1993). Point Optimal Tests for Testing the Order of Differencing in ARIMA Models. *Econometric Theory*, 9, 343–362.
- SCHICK, A. (1986). On Asymptotically Efficient Estimation in Semiparametric Models. *Annals of Statistics*, 14, 1139–1151.
- STAIGER, D., AND J.H. STOCK (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65, 557–586.
- VAN DER VAART, A. (2000). *Asymptotic Statistics* (first edition). Cambridge University Press.
- WHITE, J.S. (1958). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *The Annals of Mathematical Statistics*, 29, 1188–1197.
- ZHOU, B. (2017). A New Semiparametric Approach for Nonstandard Econometric Problems. *Working Paper*, Tilburg University.
- ZHOU, B., RAMON, VAN DEN AKKER, AND B.J.M. WERKER (2016). Semiparametrically Optimal Hybrid Rank Tests for Unit Roots. *working paper*, Tilburg University.